DOCUMENT RESUME

ED 335 372	TM 016 970
AUTHOR	Croft, Cedric, Ed.
TITLE	Conference on National Monitoring in Education, Proceedings (Wellington, New Zealand, September 1990).
INSTITUTION	New Zealand Council for Educational Research, Wellington.
SPONS AGENCY	Ministry of Education, Wellington (New Zealand).
REPORT NO	ISBN-0-908916-00-0
PUB DATE	Sep 90
NOTE	116p.
PUB TYPE	Collected Works - Conference Proceedings (021) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Conference Papers; *Educational Assessment; Educational Policy; Elementary Secondary Education; *Foreign Countries; *Government Role; *National Programs; Needs Assessment; Public Policy
IDENTIFIERS	Australia; *Monitoring; *New Zealand

ABSTRACT

The debate about national assessment in education is in its infancy in New Zealand. It is recognized that the requirements of assessment for national monitoring may conflict with assessment of a formative nature within the schools because of differing requirements for the numbers of students assessed, the type of assessment advisable, and the resources to be devoted to the task. The publication of "Tomorrow's Standards" in July 1990 provided the impetus for a conference on national monitoring that would include the views of the two major political parties. The following conference papers are provided: (1) "Introduction" (C. Croft); (2) "Overview of National Monitoring in Education" (T. Crooks); (3) "Government's Response to Tomorrow's Standards: Policies and Initiatives" (P. Goff); (4) "Symposium on Minister's Proposals" (C. Parker and others); (5) "Perspectives on National Monitoring" (D. Greig); (6) "National Monitoring in a Treaty of Waitangi Framework" (K. Mataira); (7) "The Opposition's Response to Tomorrow's Standards: Policies and Initiatives" (L. Smith); (8) "Symposium on Opposition Spokesperson's Proposals" (S. Hearn and others); (9) "Australian Developments in National Monitoring" (G. Masters); and (10) "Concluding Remarks" (I. D. Livingstone). The names of conference contributors and participants are listed, and the conference agenda is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

Proceedings of the

Conference on National Monitoring in Education

ED335372

NZCER
September, 1990

Edited by Cedric Croft

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. A. PICKENS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

New Zealand Council
for Educational Research

TM016970

BEST COPY AVAILABLE

Proceedings of the

Conference on National Monitoring in Education

NZCER
September, 1990

Edited by Cedric Croft

New Zealand Council
for Educational Research
Wellington
1990

Conference sponsored by the
New Zealand Council for Educational Research
with support from the
Ministry of Education.

©1990
New Zealand Council for Educational Research,
Box 3237,
Wellington,
New Zealand.

ISBN 0-908916-00-0

Contents

Page

Introduction - Cedric Croft 1

Overview of National Monitoring in Education - Terry Crooks 5

Government's Response to *Tomorrow's Standards* : Policies and Initiatives - The Hon. Phil Goff 22

Symposium on Minister's Proposals 30
- Carol Parker 32
- Cedric Hall 37
- Marilyn Davies 48

Perspectives on National Monitoring - David Greig 54

National Monitoring in a Treaty of Waitangi Framework - Katarina Mataira 60

The Opposition's Response to *Tomorrow's Standards* : Policies and Initiatives - The Hon. Lockwood Smith M.P. 63

Symposium on Opposition Spokesperson's Proposals 72
- Shona Hearn 74
- Ross St. George 82
- Ken Wilson 87

Australian Developments in National Monitoring - Geoffrey Masters 91

Concluding Remarks - Ian Livingstone 102

Page

Contributors 106

Conference Participants 107

Conference Programme 109

Introduction

Cedric Croft
Chief Research Officer
NZCER

Assessment practices and assessment philosophies have been perennial issues for debate in local literature during the last 20 years at least, but the publication of *Assessment for Better Learning : A Public Discussion Document* in September 1989, and *Tomorrow's Standards* in July 1990, signals a new emphasis in educational assessment. This shift from examining individual learners to assessing institutions and monitoring the effectiveness of the national system as a whole, is likely to set the agenda for assessment in the immediate future. New Zealand does not stand out as a result of these assessment initiatives, as we are now beginning to follow the path of countries such as Australia, Canada, the United States and the United Kingdom, all of which have instituted national monitoring of educational progress.

It is now recognised that the distinctive requirements of assessment for national monitoring and assessment of a formative nature within schools, have the potential to conflict. National monitoring may be undertaken by sampling students, but every student should be assessed if the intention is to focus instruction for all learners. Until the advent of Standard Assessment Tasks (SATs) as part of the approach to monitoring in the United Kingdom, national assessment tended to consist of objective paper and pencil tasks well suited to mass administration and machine scoring. Inevitably these have emphasised the more easily-measurable aspects of the curriculum and therefore ignored important curriculum outcomes more readily sampled in a formative assessment mode.

Another area where national assessment and formative classroom assessment will differ is in respect of time between the measurement being undertaken and the results becoming available in a form that may be used by schools. A further conflict will be found in the major financial, technical and human resources available to national assessment, ensuring this activity is better endowed than are individual schools or groups of schools, to produce and analyse superior measurement instruments on assessment techniques. This in turn may mean greater public confidence in national assessment procedures, despite these not having the flexibility to cater for developments of particular local significance, unless of course national monitoring is built around a set of 'anchor items', utilises an 'item bank' structure, and is to be augmented by incorporation of a meaningful local component.

A final tension between national assessment and formative assessment to be mentioned here may be found in the control of the activity. At this stage national assessment in New Zealand appears to be driven by conservative political and economic pressures. In these dual terms the types of questions that may be asked of national monitoring are: How effective is our school system? What standards are being achieved in our schools? Are these standards improving or declining over time? Is the education system delivering value for money? Where might we better direct spending on education? How might we achieve better results for less expenditure? What then should be asked is whether answers to these questions have the potential to improve the quality of education?

Even if school assessment is a component of an external examination environment, the utilization of the assessment data resides largely in the school and the assessment may be better directed at educational questions. How well has this class mastered this particular content? Where are the most obvious weaknesses in their learning of this content? What order of merit has been displayed on this content by my students? Who is now ready to progress to more complex material? How does the achievement demonstrated here compare with that of other examination classes? What changes should I make to my teaching strategies? Assessments of this nature have the potential to improve learning, national assessment has the potential to demonstrate learning. Points such as these help to illuminate the major claims for and against national assessment as they are explored further in some of the accompanying papers.

Given the potential impact national assessment might have on education and given that for New Zealand it will be a new enterprise, the time was judged to be right for a conference on the theme of national monitoring in education. NZCER has been active in the fields of examining, achievement testing and advising on school assessment for many years now and as a consequence has developed a group of staff with experience, expertise and interest in this area. Accordingly, the Board of NZCER gave their enthusiastic support to a proposal for this seminar at NZCER, as it proved an opportunity for the organisation to contribute to professional development in the assessment field.

A planning group of Jacqueline Burgon, Richard Benton, Valerie Podmore, Neil Reid and Cathy Wylie was set up under the chairmanship of Cedric Croft, to decide on the focus of the conference. Although a number of dimensions and issues were discussed it became clear very soon that *Tomorrow's Standards* and its proposals for national monitoring, would prove timely and be of intense professional interest. At the outset it was decided that on this occasion NZCER staff or Board members would not be invited to deliver papers or take part as symposium speakers, but should contribute to the organisation on the conference day by chairing sessions and being prepared to lead discussion. With the 1990 General Election in mind it was decided that the views of the two major political parties on *Tomorrow's Standards* would be important, and would provide a good backdrop to additional commentary of an educational, economic and cultural nature. This then became the central theme of the one day conference. The final programme reflected this general intention very clearly.

The opening paper by Dr Terry Crooks, who was a member of the ABLE Working Party, surveys moves towards national assessment overseas and outlines clearly the major concerns and issues of the Working Party. He also gives a clear and thoughtful summary of the Working Party's discussions on national monitoring and shows how these discussions shaped their final position. The papers delivered by the Hon. Phil Goff and the Hon. Lockwood Smith were each followed by a symposium designed to provide commentary from a broad range of perspectives, on these papers and *Tomorrow's Standards*. The views of the two politicians on assessment issues are important, as each paper shows clearly the priorities of the major parties and the broad agreement they have on the importance of monitoring the education system more formally than at present. It is worth noting here too, that the symposia and ensuing discussion provided a valuable aspect of this conference and that a brief comment on each speaker's contribution is included prior to the symposia papers.

David Greig has examined the concept of national monitoring with the eye of an economist, a perspective that is probably misunderstood by educators and regarded as a view that seeks to limit educational funding. He makes the important point that some independent appraisal of the effectiveness of government spending is necessary, irrespective of the portfolio, but points out too that appraisal by itself, is unlikely to improve efficiency. Katarina Mataira articulates perspectives on national monitoring within a Treaty of Waitangi framework and raises fair questions and concerns about the impact of this move on the pursuit of goals within Maori education. Clearly, any move that is seen as reducing autonomy is unlikely to be embraced with enthusiasm. Geoffery Masters brings to our debate Australian experiences of national monitoring and illustrates part of what he has to say by reference to recent state-wide assessment in New South Wales. Many of his comments provide valuable pointers for New Zealand's proposed developments. In his concluding remarks Ian Livingstone sums up a number of the main points covered, and asks, Does national monitoring of standards really bring about any changes in the system? He concludes on the optimistic note that our teachers will subvert the bad and enhance the good features of any forthcoming reforms in assessment.

Given that the debate about national assessment in New Zealand is in its infancy, this set of Proceedings provides a worthwhile input at this early stage. It would be a mistake to embrace national assessment too wholeheartedly, as the evidence to support its beneficial impact is virtually non-existent. On the contrary too, the prophets of doom who suggest nothing but negative effects on the curriculum, pupils and teachers also have no direct evidence on which to base their claim. At this stage a healthy scepticism regarding the possible dangers or benefits of national assessment would seem prudent. It is regrettable that there has been so little informed debate in New Zealand on the subject, but we trust these papers will contribute to the process.

Overview of National Monitoring in Education

Terry Crooks
Higher Education Development Centre
University of Otago*

I am going to spend the first ten minutes or so briefly outlining the thinking underlying the report *Tomorrow's Standards* and the work of the working party leading up to that report. I will then devote the remainder of the time to giving an overview of national and international procedures for monitoring educational performance.

Background

Why was the working party on *Assessment for Better Learning* set up? The working party understood that David Lange as Minister of Education and Noel Scott as Under-Secretary for Education saw the reforms of education that they were initiating as a two-stage process. The first stage became *Tomorrow's Schools : the reform of educational administration*. The second stage was to focus attention on the quality of teaching and learning. Our working party was to lead that second focus.

Most of you will be familiar with the terms of reference for the working party. They give primary emphasis to two things: ways to monitor the effectiveness of the New Zealand education school system on student learning, and ways to evaluate the effectiveness of the work of different schools. Nevertheless, from the very first meeting of the working party there was a firm conviction that while we had to address those issues, we were not convinced that these two issues were at the heart of better learning for students. Therefore the Working Party quite deliberately, and with a great deal of conviction and cohesion, decided to roam more widely than that, and in particular to give detailed attention to the ways in which the progress of individual students is assessed.

---

* Dr Crooks was a member of the working party on *Assessment for Better Learning*.

We discussed this plan with the politicians and officials who had set up the working party. They agreed to let us interpret the terms of reference widely, and indeed continued to give us complete freedom to follow our own judgement throughout the project. There were, however, some questions from various education sector groups about our broad interpretation of the terms of reference. In particular, there was concern about our attention to assessment and qualifications in the secondary school, and about the attention we gave to issues of teacher assessment.

There is one other general point I would like to make at the outset. This relates to the consultation processes undertaken in preparing the report. In comparison with what has occurred during some recent 'reforms', I believe we managed a better level of consultation. Before we even started to prepare our public discussion document, we invited submissions from all interested parties. We then held a seminar at Lopdell Centre, to which we invited a wide range of people with an interest in education. We asked them to guide our thinking and respond to our early proposals. With this preliminary work behind us, we prepared a public discussion document and allowed a period of three months for responses. Taking all the feedback into account, we prepared a final report which is significantly different, in important details, from what was proposed in the public discussion document. Thus we perceive the final report to be a truly collaborative product.

Central Issues in *Tomorrow's Standards*

There are several key thrusts underpining the recommendations in *Tomorrow's Standards*. I would like to run through these briefly.

First, and most importantly, our overall goal throughout was optimal development for all children in the New Zealand education system. We were not focussing on any one segment or sector of students; we were concerned for all. We were seeking excellence for the most capable students, but also success for students in categories which experience considerable educational failure. We saw the cost of neglecting to pay attention to the needs of all students as disastrous. Some of the current ills in our society, we believe, may have arisen because we have not adequately addressed the educational needs of a substantial proportion of our students.

A second important conclusion that we reached is that accountability is an important issue to be addressed in the education system. After all, the total budget for education is 4.3 billion dollars. That is not a sum that can be spent without some attention to accountability. However, the purpose of accountability always has to be seen as ensuring quality and improving quality. Thus, in addressing issues of accountability, we were very conscious that we wanted to address them in ways that actually did lead to improvement, and which had no serious negative side-effects. We believed that a key consideration was the motivational consequences, for the participants, of the accountability processes that were to be put into place. If the procedures were not motivating, did not encourage efforts to improve from the people who were involved, then the procedures would not be desirable.

A third general point is that the working party recognised and accepted that comparisons will occur: comparisons among students, comparisons among schools, comparisons among national education systems. But we were convinced that the focus of any system of assessment that was put in place should not be on comparison, it should be on growth. This implies and requires the monitoring of outcomes across time, whether for the national system, for individual schools, or for individual students.

Fourth, to get an adequate picture of what is happening, a broad range of measures and outcomes is required. You cannot get an adequate picture by focussing on certain key measures and variables, because as soon as you focus on these you run the risk that you will get a distorted picture of the overall pattern. To take a simple example, a rigid and punitive teaching method that works well in inculcating basic factual knowledge may have disastrous effects on long term motivation for learning that subject, or on the development of thinking skills. If only factual knowledge is measured, then there will be an inadequate picture to judge the effectiveness of the particular teaching approach. Even more breadth of assessment is required if we want to go beyond understanding of what is happening to understanding of why it is happening. If we truly want to be able to change what's happening in education, then we need to be able to understand why particular results are occurring.

The fifth point is to some degree implied in what I've just said. In measuring educational outcomes it is not enough to measure cognitive variables - we certainly also need to measure psychomotor variables and, perhaps most importantly of all, we need to measure variables relating to motivation and attitude. We tend, when we think of measuring educational outcomes, to look only at achievement outcomes. This gives an inadequate picture of what is happening. I'd like to share with you a couple of wonderful quotes about the importance of motivation.

> "The chief impediments to learning are not cognitive. It is not that students cannot learn, it is that they do not wish to. If educators invested a fraction of the energy that they now spend in trying to transmit information in trying to stimulate the students' enjoyment of learning, we would achieve much better results."

This is the abstract from an article on intrinsic motivation by Mihaly Csikszentmihalyi (1990). In his article, he goes on to discuss all the efforts that have gone into trying to develop new ways of presenting and teaching information: computer-assisted learning, embedded questions in textual material to foster thinking about the material, programmed learning, and so on. You could also add such local developments as Beginning School Mathematics and Reading Recovery. His comment about these efforts is extremely pithy and brief.

> "The shining knights who go forth equipped with the latest weapons of cognitive technology returned blackened and bruised, while the dragon of ignorance goes on peacefully slumbering in its cave."

The working party is convinced that motivation is a key factor that must be taken into account, both in planning assessment procedures and in measuring the outcomes of education.

Let me take just one example of a recommendation in the report, and indicate how these factors influenced it. We were asked to look at the evaluation of school effectiveness. In particular, many of you will be aware that there was quite a suggestion that the working party should focus on value-added assessment, developing ways to measure the extent to which different schools add value to their students.

The Minister of Education at the time wanted a measure that would enable Nga Tapuwae College and Auckland Grammar School to be compared in terms of the value added. He had quite a strong assumption that, in fact, Nga Tapuwae would come quite well out of the comparison.

We looked at the value-added option in considerable detail, but concluded that this was a near impossible task. There are all sorts of potential difficulties associated with assessing the value added by schools, and we found many arguments both practical and philosophical against it. One of the practical arguments is that the studies which have attempted to compare the value added by different schools have found that the ranking is quite unstable - that it is different in different years, different in different subjects, and different depending on what intake variables are used to assess the value that students brought into the school.

Beyond that, there are obviously multiple influences on individual students and on what they learn and achieve. These include their peer group and their families, as well as their school. Even if you could control for all of those influences, you would still have the difficulty that you couldn't satisfactorily compare the amount gained by students starting low on the outcome scale and students starting high on the outcome scale. In psychometrics, there are all sorts of wonderful techniques, but there is no easy way of ensuring that a gain from two to five on an educational scale is of the same merit as a gain from five to eight. The problem arises because we do not have interval measuring scales. If we had a scale where Nga Tapuwae took its students, entering in the third form at level 2, up to level 5 by the time they were in the fifth form, while over the same period Auckland Grammar took its students from level 5 up to level 8, could we really say that these two gains of three were in any sense equivalent. Almost certainly not. It might be possible to compare schools where the intakes were at about the same level when they came in. You might then be able to compare the outcomes and come to some conclusion, but it would be a very fragmented and very restricted conclusion.

Having rejected the value added proposal, the working party then looked at other ways of evaluating the effectiveness of schools. One way was clearly external review. We have a history of school inspections over many, many years in New Zealand. We also have a proposal before us that evaluation of the effectiveness of schools would rest quite heavily on visits by teams from the Education Review Office.

With external reviews, no matter how hard the evaluation teams try to inculcate a sense of co-operation and of helpful assistance into the process, there is to some degree likely to be an adversary role in the evaluation process, a temptation to hide key information in order to present a better picture than might otherwise be the case. Also, we were concerned about the difficulty of adequately involving the community in an external review, in a situation where a team came into each school for a comparatively brief period of time.

We concluded that the appropriate recommendation in the area of evaluation of school effectiveness was that the central process be a process of self-evaluation. This is not a soft option. It demands a lot of the school personnel and the school community. But it has several advantages. The school is able to focus on its own goals and its own growth. It can involve the community heavily in the process. It is more likely to lead to change because the people involved are participating and making their own commitments as they go, starting with the choice of what will be evaluated. Key things that are concerns about the school will be addressed.

There will still be a valuable role for the Education Review Office, but now it becomes a role in which staff come in and react to the self-evaluation that has already occurred. They can bring an external perspective, having looked at quite a range of schools. They also can bring in ideas about how to achieve some of the goals that the school has set itself, but which it is struggling to achieve. Of course, they may also bring in a perspective which says "You should have looked at this. You've chosen to put the blinkers on and look at the things that you wanted to focus on, but the Government wants you also to focus on these other things. We suggest that the next time you review, you have a closer look at X, because we believe it's important and the Government has said that you must address that area."

This example illustrates the complex interplay between the factors which I introduced earlier: balancing accountability with formative evaluation, emphasising growth rather than comparison, using a wide range of measures, and paying special attention to motivational considerations.

National and International Monitoring of Education

I would like now to move to the main thrust of my talk this morning, which is to overview national and international monitoring at the level of the education system. There are two thrusts to my comments. The first is to talk briefly about educational indicators. The second is to devote considerably more time to procedures for assessing student outcomes. The reason for devoting more time to the assessment of outcomes is that this is an area which has not been addressed adequately in the New Zealand education system.

Educational indicators are nothing new. They've been used for many years in New Zealand. Most of you will be familiar with the annual publication of the Ministry of Education which presents *Education Statistics of New Zealand*. It is a book full of educational indicators, about what's happening in our education system. We've also had international tables of educational participation put out by the OECD for many years.

What do educational indicators look at? They look at the resources of education - teachers, curriculum materials, facilities, finance: they look at participation in education of different age groups; they look at the initial status of learners; they look at the organisation and processes of education; and they look at the outcomes of education, such as the achievements and attitudes of learners and the longer term consequences that flow on from participation in the education system.

When selecting educational indicators, it is vital that we capture the important qualities. There's a great danger of measuring only the easy things. Some things are easy to measure, capture and record. The danger is, of course, that if we record the easy things and miss other important things, the picture that is presented and the interpretations we make will be misleading.

Some of you will have been involved in looking at the proposed indicators for the tertiary sector. A major question in most tertiary institutions has been "Do they really capture the guts of what we're trying to achieve?" That is the key question to ask about any system of educational indicators. If it doesn't capture the 'guts' of what we are trying to achieve, then we will have to provide other mechanisms for incorporating that information.

It is also important to measure things that are of local importance. When we are using international systems of educational indicators, we are bound to miss things that matter to New Zealand. When we are talking about using a national system of educational indicators, we are bound to miss things that matter to individual schools. Therefore it is important to allow, in any system of educational indicators, for locally important things to be included. In New Zealand, for instance, some of the requirements in the charter framework clearly imply a need for local educational indicators, indicators which would not make a lot of sense for the whole OECD.

The final point I wish to make about indicators is that the working party believes a great deal of attention needs to be given to interpreting indicators in context. Indicators should be interpreted in relation to the particular context in which they've been collected. There are local patterns which can affect the interpretation. Let me give you a little example which originates with Joan Paske, an example of a medical diagnosis process. If a doctor checked me out right now, it would emerge that I have a raised temperature, pulse rate, and blood pressure. There are lots of potential diseases or illnesses that these indicators could be signs or symptoms of. It requires a knowledge of the context in which the measures were obtained to recognise that these results are not the consequence of illness, but just of the stressful position I'm in right now!

Let me now move to assessing student outcomes, because that is a major thrust of today's session. The working party believes that procedures for assessing student outcomes is an area where our national monitoring is weakest. We have well-established measures of quantity of education, but we have very weak measures of key components of the quality of education. We have good measures of participation and attendance, but our measures of achievement, motivation and attitudes are very weak, for the purposes of national monitoring.

Our national examinations and national awards in the secondary school are quite inappropriate measures of national quality. They have nothing to do with a fixed standard. Just to briefly illustrate this, if every student in the fifth form at the start of next year were encouraged by their teachers to go off to the beach for the first term, to the ski slopes for the second term, and back to the beach for the third term, it wouldn't make the slightest difference to the national distribution of School Certificate results. It would affect individuals' results, but it wouldn't change the national distribution, because of the scaling and moderation procedures that are adopted. Sixth Form Certificate results are based on the fifth form results, so they are in exactly the same boat. A similar picture applies at Bursary level: fixed percentages of the candidates are awarded A bursaries, B bursaries and University Entrance passes. Over the last fifteen years, the proportion of entering third formers sitting Bursary has doubled, but at the same time the proportion getting A's has remained the same. It seems most unlikely that the real standard for an A bursary has remained constant.

Standardised tests are also used as means for assessing student outcomes. A danger with the use of standardised tests for national monitoring is that item relevance decays over time. Some items which made sense ten years ago no longer make sense now. The norms decay across time, vividly illustrated by the recent evidence in the United States that every State in the country is performing above national norms on some standardised tests. The reason for this appears to be that the norms are rather old, and there has been a national improvement in the average scores on these tests. Few of these standardised tests have procedures for maintaining standards across time. One that does is the *Scholastic Aptitude Test,* where there has been evidence provided for quite some time of shifts and trends in the scores.

As a result of such considerations there have been developed, over quite a number of years now, systematic national and international monitoring projects looking at student achievement. My comments on these projects are based not on a thorough survey of such activities, but on examples which have come to my attention.

The National Foundation for Educational Research in England conducted reading surveys between 1948 and 1964, to look at the standards and performance in reading in British schools. Scotland conducted national surveys in 1953 and 1963 in key core subject areas - literacy and numeracy. The International Association for the Evaluation of Educational Achievement (IEA), set up in the mid-1960's, has been conducting surveys for a number of years, and New Zealand has participated in some of these. *Reading Comprehension in Fifteen Countries* was the report from one of the very early surveys, and New Zealand also participated in the mathematics survey at the beginning of the 1980's. Fifteen or more countries participated in each of these surveys, and New Zealand now has a commitment to regular participation in future surveys of the IEA.

The National Assessment of Educational Progress in the United States, set up in the late 1960's, has been surveying achievement at certain grade levels and in quite a range of school curriculum areas (and even beyond the school, into adult literacy). The Assessment of Performance Unit in the United Kingdom, set up in the early 1970's, has been conducting surveys since then. There are provincial surveys in almost every province in Canada, and Geoff Masters will bring you up-to-date with what's happening in Australia, where most of the states have some sort of monitoring project underway. Finally, I note that the new assessment provisions in England and Wales are not only capable of providing results to be used for national monitoring, but also are going to be used to publish results about individual schools and to provide data for individual pupils.

There is a pattern that has emerged from most of these surveys. First, they tend to focus on certain core subjects. Some of them have solely these subjects as their focus, while others say that the core of the programme involves surveys of two or three subject areas, but surveys of other subjects are added in from time to time. There is typically quite wide consultation with professionals about the objectives to be assessed. In almost all of the survey programmes, multiple matrix sampling or some other sampling processes have been used so that different students are tested on different sets of tasks. This means that the results are only amenable to being looked at at the national level, and cannot be used for producing data on schools or individuals. Typically, the same subject is reassessed after several years, so that it is possible to get a picture of progress across time. This requires an attempt to preserve a significant proportion of the items from publicity, so that they can be used in the next round of testing, and then the results compared from one occasion to the next.

The Project ABLE team looked at possibilities for national monitoring of outcomes in New Zealand. There were quite a number of issues that we had to address. I would like to talk about each of those issues in general, and then briefly say what we concluded about them.

First, it is necessary to decide what levels of schooling should be monitored in these programmes. If you look at what has happened internationally, you see a pretty fair consensus. What tends to happen is that the organisers pick the last level at which most students are participating in the education system. They also tend to pick the earliest level at which students can handle testing reasonably independently without individual testing. Finally, they often pick a point about halfway in between. In our case, we chose Standard 2, Form 2 and Form 4, although the form of the Form 4 testing was to be rather different from the other two.

Second, a decision must be made about whether to sample intact classes or individual students. Sampling intact classes is obviously most convenient: it is much less disruptive for the work of the school and it saves a considerable amount of effort. Sampling individual students allows fewer students to be tested to get the same degree of accuracy of national results, but at considerably greater organisational cost. The ABLE working party recommended the use of intact classes.

One of the most important issues to be addressed is the range of curriculum objectives to be monitored. The view of the ABLE working party was that it was not satisfactory to select certain key curriculum areas and focus on these. We believed very strongly that this would exert a distorting effect, both in the picture that would emerge and in the longer term impact of how schools would adapt to the monitoring that took place. We believe it is essential that the sampling range across all the objectives that can conceivably be held for the education system. For instance, objectives that might be primarily of interest to the Maori community and of very little interest to the Pakeha community should be included.

We also came to the conclusion, and I think that this is a very important conclusion, that we should not be focussing only on objectives that are associated with the curriculum at the level that is being tested. In other words, in the Standard 2 monitoring we should not be looking only at the objectives of the Standard 2 curriculum and testing these, because that is not going to give an adequate picture. It would also increase the risk that the monitoring would become a powerful influence on the teaching and learning that takes place at that level. Instead, the assessment should sample from objectives that most Standard 2 pupils will have thoroughly under control, all the way up to objectives that few Standard 2 pupils will have even attempted. We are trying to look at a snapshot of what students can do, rather than an analysis of how students perform against the curriculum of a particular level in the school.

It is necessary to decide how much emphasis there should be, within the cognitive domain, on factual learning compared with deeper skills. One of the difficulties is that in reporting the results to the public, it is very attractive to the media and to the public to report on factual learning. If you think about recent reports of the survey of science achievement that is being undertaken in New Zealand (to compare with similar surveys in British and the United States), the things that were attractive to report related to what proportion of New Zealand students knew fact X versus the proportions of American and British students who knew fact X. It is much harder to report clearly on results relating to deeper skills, such as ability to present coherent written arguments.

However, the working party emphasised strongly the importance of assessing the deeper skills, rather than concentrating on the facts that have been learned. We wanted as far as possible to give schools flexibility about what they teach, so that they can select the specific content they use to teach the underlying skills.

To illustrate this, one of the benchmarks in my own thinking has always been if I was involved in teaching sixth form Geography, I would want an assessment system that allowed or even encouraged me, in the event that the Abbotsford landslide occurred in my city, to take my students to look at that vivid, important event. This would be something that they would never forget for the rest of their lives. If the assessment system actually disadvantaged the students because we took two weeks out to focus on that event, and on the Geology behind it, then there would be something wrong with the assessment system. These sorts of considerations argue against using measures of factual learning as the prime measures in monitoring educational outcomes.

Another issue to address is whether the testing should allow the students plenty of time, or should be speeded. Obviously there are certain areas where speed will be important. For example, if we are including in our Physical Education testing an assessment of how quickly students can run fifty metres, then that's a speeded activity. The working party does not believe, however, that it is appropriate for most of the testing to have rigid time limits, so that students are struggling to complete it. Most of the work that we do in our community is not subject to the sort of time constraints that we often impose in our assessment processes.

If we are going to use sampling procedures, such as multiple matrix sampling, we must decide whether to overlap the samples of items or to have different students attempt completely different sets of items. If completely different sets are used, then the results for one set of classes are completely distinct from the results of another set of classes, because they have been tested on completely different material. The disadvantage of this is that it is not possible to calculate correlations among all the sets of items, nor to examine for the whole sample relationships between teaching methods, classroom activities and the measured outcomes. To allow for this the IEA Maths Survey had a core set of forty items that were used in all classes. Then, in addition, there were four different sets of items that were randomly distributed among the students. In the United States, the National Assessment of Educational Progress has developed what they call a balanced incomplete block design, where different blocks of items overlap in a carefully planned fashion so that researchers can look at inter-relationships. The ABLE working party did not spend much time on this issue, seeing it as a technical detail to be addressed later.

Another issue is how long the interval between surveys should be. The shorter the interval, the less likely it is that you will be able to identify change. On the other hand, the longer the interval, the harder it will be to maintain a core of items that can appropriately be used on both occasions. Educational objectives change. To take a simple example, the sorts of words that you can include in a reading comprehension test change with time. A word like 'wheelwright' made a lot more sense to students thirty or forty years ago than it does now. If you had such a word in your monitoring the first time you did the testing, and you did not retest until ten years later, the word might not be appropriate then. The general consensus is that if you want to be able to see and measure change, an interval of about four to five years is appropriate. That is what the Working Party suggested.

One of the real difficulties, where you're attempting to assess change across time, is how you can talk about the results of one survey without revealing all of the items. If you do so, the next time you repeat the monitoring some of the items may have been incorporated deliberately and specifically into the teaching, distorting the findings. Thus there is a difficulty, because you cannot describe the results very clearly without referring to examples of the items; and yet if you reveal all of the items you are in trouble for the next testing. Most of the monitoring projects have developed a guideline that some of the items be revealed to facilitate public reporting, but that some not be revealed, to allow satisfactory reuse.

There are issues about how much to aggregate the data. Some of the reporting has been done on the basis of total subject performance, such as performance in reading or in mathematics. In other cases, it has been done on the basis of performance in individual items, or in clusters of related items. The working party, because its focus was on how the result could be used to improve education, saw the importance of focusing on the finer detail, rather than on broad summary measures. Indeed, we would recommend strongly against reporting an overall summary figure for reading or mathematics.

It is also important to decide how much emphasis should be placed on contextual data. In the New Zealand IEA mathematics study, it was found that 92 percent of the low-gain classes in the third form level were streamed and 63 percent of the high-gain classes were streamed. Because that is just a correlational finding, causality cannot be inferred, but the finding certainly tends to negate the often-expressed argument that streaming creates more efficient learning conditions. In this study, the more streamed environments tended to perform less well than the unstreamed environments. Another interesting finding was that the teachers teaching the high-gain classes had an average of six years of teaching at that level, whereas the teachers of the low-gain classes had an average of three years teaching mathematics at that level. Again, you cannot directly infer causality. It may well be that the reason people have had six years teaching at that level is because they have been identified as teaching particularly well at that level. Nevertheless, these interesting pieces of data allow some interpretation, and suggest some possible actions. The availability of contextual data about teachers and school organisation, in addition to the outcome data, made it possible to undertake such analysis.

Finally, we come to the issue of how to use the data as a basis for educational decision-making. That is a really difficult issue. The working party believe it would be important to present results on the percentages getting items right and on the sorts of errors that were being made, so that we could understand what sorts of misunderstandings and misinterpretations were occurring. It could be valuable to aggregate small groups of items to present summaries of achievement in various skill areas (but not to aggregate across broad domains). Perhaps Geoff will talk later in the day about some of the work being done in Australia to define performance bands, based on item response theory. Another stage of decision-making would involve examining the interrelationships of variables, to try to understand what is happening.

The working party believes that the use of the results is going to depend heavily on human judgement. For example, members of the Policy Division in the Ministry, in consultation with teacher groups and others, would look at the results and decide they were not satisfied with student performance in some areas, and that significant action was needed. They might also note a particular relationship which was observed, and ask whether the results implied that more resources should be provided in a targeted way, to raise achievement in a particular area.

It would also be important to consider the issue of how much and in what form the results for different sub-groups in the community should be reported. Should we report results for Maori students, for rural students, for males, for females, etc., and how might those results be used to foster understanding of what is going on and lead to a greater degree of equity.

To sum up, it was the view of the ABLE working party that improved monitoring of national education outcomes is possible, is not unduly expensive, and can be done in ways that have minimal negative side effects. However, we did not believe that national monitoring initiatives were crucial to educational development in New Zealand. We gave higher priority to some of our other recommendations.

References

Mihaly Csikszentmihayli (1990), Literacy and Intrinsic Motivation, *Daedalus*, September, pp. 115-140.

Government's Response to *Tomorrow's Standards*: Policies and Initiatives

The Hon. Phil Goff
Minister of Education*

Thank you for the invitation to speak to your conference this morning. Your invitation sets out a variety of issues in relation to National Monitoring in Education, from the considerations which led to the establishment of Project ABLE to the steps necessary to implement its recommendations.

The working party report originated from a concern about standards in New Zealand education. There was a recognised need for a more effective means of monitoring New Zealand's educational standards. We needed improved methods of assessing how effective schools are in developing student learning. We needed an improved method of assessing the progress of students throughout the school system. Parents wanted to know how well their children were learning and how effective their local schools were.

There is a propensity for the wider public to believe that standards are slipping in New Zealand. Interestingly, research shows that it is an apparently common perception by the public about education, right across the western world. To provide the answers parents want and to confront perceptions of slipping standards, we needed a clearer and more factual basis for measuring movement in standards in education over time in New Zealand and a means of comparing how New Zealand education shapes up alongside other countries.

It has been a weakness in the New Zealand education system that we have lacked sufficiently rigorous and comprehensive procedures and practices, and the means of reporting these to parents and the wider community, to enable the effectiveness of the system to be judged. Precise, accurate and reliable information provided by systematic monitoring systems, strategies and reporting procedures is needed to answer questions put by the general public, business community, teachers and politicians.

---

* Since delivering this paper Mr Goff has not retained his seat in Parliament.

It is also needed to assist the development of policy advice and to illustrate more clearly the reasons underlying policy decisions. Most of all however we need better monitoring procedures in education in order to improve our performance and enhance the learning environment. Where there are problems we need to identify them and take steps to remedy them.

The title given to the working party we set up was deliberately chosen. It was a working party set up not simply for the sake of assessment, but rather for *Assessment for Better Learning*. The purpose of monitoring the performance of the education system must be to enable the findings of that monitoring to be used to enhance the system.

The report on Assessment for Better Learning, *Tomorrow's Standards*, had its origins in a meeting in March 1989. The decision was taken to establish a working party to fully investigate monitoring of education performance. Terms of Reference were drawn up which asked the working party to recommend to Government procedures for assessment to achieve two purposes. Firstly, they should monitor the effectiveness of the New Zealand school system on student learning. Secondly, they should enable assessment of the effect of individual schools on students' learning achievements.

The composition of the working party reflected a broad range of expertise and experience in diverse areas of education which I believe made it one of the strongest and most effective working parties established to conduct any educational enquiry. I want to thank Sir James Stewart, Yvonne Brown, Dr Terry Crooks, Dr Geraldine McDonald, Hana Tukukino, Rae Munro and Joan Paske for their contribution. I also want to acknowledge the role played in establishing the project by Noel Scott, who has had an active interest in this area over a long time.

Between April and June 1989, the working party identified issues and explored a range of possible procedures for the assessment of students' learning. They studied overseas literature on assessment, visited schools, colleges and universities, and consulted widely. An initial discussion document was prepared for a further conference, approved by the Minister, and held 26-30 June 1989. The aim of the conference, attended by approximately 30 people representing a wide range of interest groups, was to provide a broader community response to the issues raised by the ABLE Project. The working party's hope was to encourage and facilitate public debate on the assessment issues, and also to ensure that the resulting recommendations would have been well canvassed and have general public support.

The public discussion document completed in mid-September 1989 outlined issues, suggested directions and asked questions. Public submissions on the discussion paper were called for by Government, and then collated and considered by the working party in developing its final report. Of the approximately 600 responses, some 80-90% showed major agreement and support for the Report's proposals in general terms, although many expressed reservations about single issues. This response was confirmed by the experience of working party group members, who had all been speaking to groups around the country over the period of the summer school holidays.

Among the issues raised were the future of School Certificate, concern over the increased teacher workloads which could occur, the possibility that education could become limited to 'teaching to the tests', and a feeling that with the rapid changes in education administration the introduction of changes to assessment should not be hurried. The relatively low level of response from individuals or groups identifying as Maori was seen as a concern and additional hui were held on marae by some of the working party members to try to address this. The submissions were considered by the working party and their response incorporated in their final report *Tomorrow's Standards*.

Assessment is an area where there are inevitably divergent opinions. I believe, however, that there will be general support for the main thrust of the working party's recommendations.

The report sets out key recommendations for assessing the system and the effectiveness of schools. Assessing the achievements of a sample of children in Standard 2 and Form 2 and a comprehensive assessment of the basic skills of all fourth form pupils is proposed. Promoting self-assessment by schools of their progress towards meeting charter goals, external assessment of schools by the Education Review Office, and the development of effective education performance indicators by the Ministry of Education, are further major recommendations. In relation to assessing the performance and progress of students at school, the report advocates a more systematic and comprehensive programme of monitoring students' achievements.

The recommendations build on effective practices already being pursued in many schools. At a primary level, that will include comprehensive measurement of progress in Reading and Writing at age 6. At ages 7, 9 and 12 it involves thorough review of each child's educational and social development by teachers and family members together. A folio of each child's work will be kept and updated, and a new system of record-keeping on each child's achievement which is now in the process of being implemented, is fully endorsed.

At the secondary level, a national basic skills testing programme will be implemented at a fourth form level. School Certificate will be retained but reassessed in two years' time. Much more flexibility will be allowed for students to engage in multi-level studies and to enter national awards at any level of secondary schooling. This will help extend those who are high achievers.

A 'Record of Achievement' for pupils, currently being trialled in nine secondary schools, will be progressively introduced across the board. This will replace the current practice of writing references for pupils, providing instead a more detailed record of skills, achievements and personal qualities. This will be of much greater value to both the students themselves and parents and employers wanting to know what the students have achieved and what they are capable of doing. A National Certificate of Education listing cumulative achievements in various national awards, which can be updated at any time, will be introduced.

Other recommendations in the report related to assessing Maori achievement, the establishment of a working party to provide guidelines on the appraisal and professional development of teachers, and strengthening pre-service and inservice training of teachers in assessment techniques. The broad thrust of the working party's recommendations are accepted in principle by Government and over the next three years we will move progressively to implement them. The strength of the ABLE recommendations on national monitoring is that they focus on the education process as well as outcomes.

As summed up in *Tomorrow's Standards*: "there is little point in monitoring the performance of the education system if the findings are not used to enhance the system. The goal is not just assessment but assessment for better learning. An approach which focuses too narrowly on attainment targets or standardised assessment tasks for all children carries with it the very real danger that it will end up driving the curriculum.

The negative effects of this are demonstrated from the United States' experience where a narrow 'testing' focus is reported as having resulted in a lowering of education standards, a narrowing of the curriculum and a strong tendency for teachers to teach to the test and to systematically coach their students on how to pass the tests. It is not surprising that many visiting U.S. educators have strongly endorsed the broadly based teaching and learning practices which they observe in New Zealand schools. There is a natural tendency in organisations for measured activities to crowd out or replace the unmeasured. The greater the consequences or outcomes from the measurement, the greater is the likelihood of other activities being displaced. A common cry in America now is "too many tests and not enough education", and there is a growing movement away from the over-reliance on standardised tests.

In England the recent introduction of a programme of standardised assessment tasks for all students at ages 7, 11 and 14, has been criticised as placing too high a burden on teachers and subverting the teaching and learning process. The system has been acknowledged by authorities and criticised as being too time consuming and costly. There is little evidence to date that better teaching and learning will be the outcome. Because of the high costs of developing and implementing national standardised assessment, there is a strong temptation to focus on subjects included in the assessment programme while others are seen as peripheral and deserving less attention.

The strength of the ABLE recommendation for light sampling is that it enables systematic monitoring of education standards without placing an intolerable burden on children, teachers or the system as a whole. A further strength of the recommendations in *Tomorrow's Standards* is that it builds on effective and proven assessment practices that are already well established in New Zealand schools. The emphasis for the regular monitoring of students' performance is placed where it belongs - with the teacher as part of the normal teaching and learning process.

New Zealand is in a good position to learn from international experience and thus avoid some of the pitfalls of over-reliance on tests that are too narrowly conceived, trivialise the education process and end up 'driving the system'. The recommendations of *Tomorrow's Standards* avoid this danger. Other positive outcomes of the recommendations are that they build on the best of what is already in place, or being developed in schools, and incorporate this into a coordinated system of evaluation which takes account of the new administration and policy structures.

The development and use of appropriate assessment tools within a coordinated system will identify areas of the curriculum where we are not achieving the desired success, and where individuals or groups are not making the expected learning gains. This will enable us to target such areas for further study, curriculum development, resources, personnel, or increased teacher training. It will also identify those areas where we are successful, help us to better understand the teaching-learning process, and to make systematic use of this knowledge to improve teaching and learning in other areas. For primary schools, it will assist teachers to further develop specific learning goals, using the aims and objectives of national syllabuses and the learning goals provided in the primary schools' record scheme. It will extend the evaluation of students' progress and improve teachers' skills to ensure that no child misses receiving the help needed to avoid failure because he or she 'slipped through a crack' in the system.

The move away from norm-referenced assessment procedures towards achievement-based, criterion-referenced assessment, already well underway in our senior secondary schools, will be further developed and strengthened. It provides Boards of Trustees with a more systematic way of effectively informing parents, whanau and community about what forms of assessment are being used and why, how their children are learning, and how well the school is meeting its charter goals. From this, it follows that, where a problem is identified, teachers will be in a better position to do something about it and parents in a better position to judge whether schools are being successful in this.

The proposals will ensure a teaching service more informed about the use of assessment measures and better able to make effective and appropriate use of them to aid teaching decisions, assist learning, and to provide soundly based, accurate information on students' progress. Objective, system-wide, over-time information on educational standards will be provided, which should counteract the present use of hearsay, rumour and myth. And, most importantly, through the procedures and techniques advocated, teaching and learning will be strengthened.

The development of reliable, valid and broadbased assessment measures and educational indicators for New Zealand, appropriate for an age where there is a demand for high level, generic skills, will take time and draw on the skills, knowledge and experience of a wide range of people. The work will undoubtedly be assisted by New Zealand's participation in international networks such as the OECD* international educational indicators project, where our involvement is with projects on "student outcomes", "education and labour market destinations", and two technical groups on "student flows" and "costs and resources", the IEA* and ACAP.*

---

*OECD - Organisation for Economic Co-operation and Development.

A short-life secretariat is being established within the Ministry to further work through the details of the proposals, and there will be close consultation with Boards of Trustees, teacher professional groups, the New Zealand Qualifications Authority and other interested parties.

I look forward to this conference contributing to that process and to the overall progress within New Zealand towards a dynamic and responsive education system.

---

*IEA - International Evaluation Association.

*ACAP - Australasian Co-operative Assessment Program.

Symposium on Minister's Proposals

Each contributor to this symposium was invited to perform two major tasks. First, they were to comment on *Tomorrow's Standards* and draw attention to features of the report that had particular implications for their sector group. Second, they were asked to discuss particular points from the Minister's speech they thought required further illustration. This latter aspect was always going to be a demanding task, given that the Minister's address was not available before the conference began at 9.00 a.m.

Each respondent made important points reflecting the perspective of their particular sector. Carol Parker reminds us of the importance of maintaining a close link between curriculum and assessment and of the potential dilemma created when the requirements of diagnostic testing and national monitoring must be reconciled. She also raises the question of the impact monitoring might have on those classroom processes not amenable to objective assessment. The issue of possible cultural bias in the *Progressive Achievement Tests* is raised also, but not explored fully or resolved to any extent.

Cedric Hall has approached his contribution from a strong theoretical and technical perspective. He discusses the essential differences between criterion-referenced assessment and norm-referenced assessment and outlines why he sees criterion-referenced assessment and externally administered public examinations as incompatible. The corollary to this is that criterion-referenced assessment and internal assessment by teachers are compatible. The paper draws attention also to the nature and impact of the margin of error in educational assessment and discusses the possible impact of reference tests on Sixth Form Certificate assessments.

Marilyn Davies' focus is on senior secondary school assessments and the proposed Records of Achievement. She makes the important point that what is taught is less important than how it is taught and learned, but acknowledges too that there are groups of employers who continue to put their faith in traditional examinations and external assessment on the grounds that these alone are fair and unbiased. When this perception is laid alongside the new basics or the Four C's of Communication, Computation, Critical thinking and Computer literacy, the range of roles employers see for assessment becomes apparent. One of the important features of this paper is that it represents to an audience of educators, the broad view of employers on some major outcomes of schooling and assessment.

Carol Parker
National President
New Zealand Educational Institute

Kia Ora Koutou,

Thank you for the opportunity to speak and share in this debate. NZEI has a critical role to play in ensuring that classroom teachers in primary schools are heard. We agree that evaluation is central to teaching and learning.

NZEI welcomes the working party's recommendations on national monitoring as being the most educationally sound given the scope of the task. In my speech I want to address briefly the dilemma of a national testing programme on the one hand and diagnostic testing on the other. This will be followed by a few caveats about testing *per se*.

The working party says:

"There are at least four reasons for undertaking regular national monitoring of performance levels in education:

- to promote public accountability of the education system as a whole, by providing trustworthy information;
- to check on the extent to which the national requirements in school charters (e.g. equity goals) are being achieved;
- to allow trends in performance to be identified;
- to help policy makers identify aspects which need special attention."

There is the dilemma I see. On the one hand information is required so that decisions can be made about curriculum policy issues and resource allocation, and on the other, information is required to determine an individual child's learning needs. Policy makers and educators must not confuse the purpose of the collection of this information and allow one to lead to the other.

It is quite interesting, and also sad, that all of this discussion and planning of evaluation has taken place before discussion or release of the National Curriculum Objectives, even at a draft stage. The discussion about assessment implies that there is a connection between assessment and effective schooling. However, assessment on its own will not help children succeed within the system. Assessment must be closely integrated with the curriculum, where it fosters learning, improves teaching, and provides valid information about what has been achieved so that sensible decision making can be facilitated.

Discussions continue on the real value and place of monitoring and assessment. In Atlanta, Georgia, at an International Reading Association Conference - members said, "This state has had, every 2 years since 1980, tests done on 4th, 8th and 12th graders, on a school by school, district by district, province by province basis."

"The test scores are not meant to be used as a basis for comparison but as a basis for improvement. Making comparisons is a cause for concern. Does the test now reflect the latest and best research. No. The existing test can be considered flawed and regressive rather than future oriented. The tests cost conservatively $8M." Atlanta's testing programme would equate to NZ's national monitoring.

We acknowledge that while education systems are different some of the problems associated with testing are universal.

What I'd like to highlight here, is that, national testing cannot access actual classroom dynamics, or diagnose why particular students have difficulty with some types of learning, or, what to do about them. Indeed, national assessment scores can be misused because they fail to account for the situation, and the teaching and learning conditions.

The second part of the dilemma is that diagnostic testing is not without faults. We have learned that standardised diagnostic testing is not value free and can be enormously damaging to different groups of people. Some schools continue to use the Progressive Achievement Tests every year when it has been proved that for Maori children in particular, the results of those tests are seriously skewed. There may be some talk of improving this testing programme and of necessary training for teachers but there is a definite cultural and gender bias. However, the '6 year net' information is an example of a national test that must continue. Diagnostic testing that is teacher-based and has at its heart the ongoing monitoring of a child's progress should be built upon. We have to learn to trust a teacher's professional judgement, and teachers must learn to share their knowledge.

In my school we keep these scores for several years, do a term by term analysis of the six year nets - note any trends that show a lack of knowledge in a particular area, or note any particular strength. We build into the junior syndicate programme, modules that will overcome the weaknesses and build on the strengths. This highlights the value of ongoing monitoring for improved learning. The idea of folios for samples of children's work is strongly supported. The primary progress records system is the link between assessment, learning and good teaching practice.

Now I want to share a few thoughts about testing. We must be conscious that much testing consists of a series of minutely sequenced, highly abstract incremental skills. The skills to be acquired become objective goals, processes and results. Yet it is precisely this approach to learning that is being rapidly discarded in today's society, because it can no longer serve the interests of production based on advanced technology. We need to guard against imagination, curiosity and invention being destroyed by the fragmentation of knowledge.

As John Boyens said at the Principals Federation Conference earlier this year, "We need to highlight 3 more R's in our 1990 curriculum - those of research, reasoning and recording, the long term whole strategies in the learning programme." I agree with him that those are crucial learning and teaching areas and they should be highlighted. We must ensure that a monitoring programme is not misused by failing to account for the conditions in teaching and learning, and by turning a learning programme from the holistic to the compartmentalized.

Educators are not opposed to measuring progress but, consistently using a core of repeated test questions which leaves only a small portion of the test available for change and improvement is not appropriate, nor is it helpful to the everyday classroom situation. There is a danger, that teachers whose class has been included in the national sample, will be held accountable for the school's results, so that the teachers see no alternative, but to slant their programmes or 'teach to the test', so there is a potential for the ambience of the school to be driven by a testing programme. That will hardly achieve teachers' and parents' aims of improved teaching and learning programmes.

The working party on Project ABLE acknowledged that cost considerations are critical.

NZEI has two major concerns about the costs involved in monitoring national performance. First, the Minister has said that more than a million dollars for the establishment stages will come from within the existing education vote. The question must be asked, "how much money do we wish to divert to a testing programme?" The costs, and benefits, of any proposed form of assessment must be weighed and continually reassessed.

It may be more appropriate for that funding to be used to improve staffing and instruction in the classroom. We must be insistent, that learning and teaching still has priority over the testing programme and that assessment simply serves these programmes. As a farmer friend of mine says fairly succinctly, "The lambs don't grow heavier just because you weigh them a lot!" Some teachers in England are spending 35% of what was teaching time carrying out tests. Secondly there is no indication of the ongoing costs.

A national monitoring programme, in NZEI's view, must be flexible enough to meet needs, and therefore, open to change and modification. For this to happen extra funding will continue to be needed, for assessment in areas that are appropriate. Monitoring must not be static. Therefore it means the ongoing costs will continue and probably grow.

We must be alert to the issues involved with monitoring the national performance, and with any testing programme. We must be cautious of the procedures used, and information gathered, to ensure that we do not progress unwittingly down another unintended path. External control over assessment, especially inappropriate use of large-scale assessment data, is recognised as a threat to the work of teachers, and ultimately the welfare and learning capabilities of our children.

In conclusion I would like to make a brief comment on the key elements of *Tomorrow's Standards*, which NZEI has welcomed. These were:

1. the proposal to build on the best things already in place, and this means existing assessment practices in primary schools should be developed, to ensure that they are comprehensive and systematic. In particular the primary progress cards should be properly implemented, with full resourcing;
2. affirming the professional role of the teacher in assessment;
3. identifying the needs for relevant training programmes;
4. insisting, that the model proposed must fit the New Zealand system and recognising that a realistic and sensitive approach must be taken to under-achievement by Maori children;
5. and finally, that for assessment to be educationally sound, it must be linked to learning programmes.

Cedric Hall
University Teaching Development Centre
Victoria University of Wellington

Tena koutou, tena koutou, tena koutou katoa,

Assessment for better learning: I liken my preparation of a response to the Minister's speech to preparation for a traditional closed-book examination of the sort... "answer any three questions from six". In the same way that a student tries to pick or spot questions to answer in such an examination, I have had to make guesses as to what might be included in the Minister's speech and gamble accordingly. I have, of course, studied the "text", *Tomorrow's Standards*, and I have searched recent Ministerial press releases - like a student studying old examination papers - for clues as to what to focus on. The analogy continues: candidates in an examination usually write down what they have prepared even if the paper asks for something different; today, you will hear what I have prepared rather than my reactions specifically to the Minister's speech.

I am going to focus on the following statements contained in the Minister's press release of 9 August and embodied in *Tomorrow's Standards*.

(i) Assessment needs to focus on measuring actual achievement rather than simply comparing pupils.

(ii) Assessment procedures need to apply consistent standards across the country.

The first statement focuses on the distinction between norm-referenced assessment and criterion-referenced assessment. More and more, educationists are stressing the importance of criterion-referenced assessment for promoting better learning. I want to look at this view in the light of the belief of many people that externally conducted examinations (as distinct from the internal assessments of pupil performance made by teachers) are essential if standards of learning are to be maintained. My contention is that criterion-referenced assessment and external examinations are incompatible: you can't have both.

The second statement interests me in the light of what we know about the accuracy of educational measurements. My view is that we should adopt monitoring procedures which are relatively effective but we should not become obsessed with precise comparability. It simply cannot be achieved. However, I shall point out what I see as a deficiency in the model proposed in *Tomorrow's Standards* (Appendix 4) for the moderation of Sixth Form Certificate assessments; in my view, this model falls short of what is needed by being too deterministic in allocating grades to schools.

Before proceeding on these concerns, I think it is important that I should declare my position on *Tomorrow's Standards*. In short, I am positive towards this report even though it was founded on terms of reference which were poorly stated and which even confused basic terms such as "assessment" and "evaluation". However, early on in their work, the Project ABLE team identified a set of sound assessment principles and followed these through into sensible recommendations for future policy. Clearly, there will be substantial costs in putting *Tomorrow's Standards* into practice and there will be a number of problems that will need to be worked through and resolved.

Criterion Referenced Assessment and Externally Conducted Examinations

> *Assessment needs to focus on measuring actual achievement rather than simply comparing pupils.*

I think we first need to understand in general terms the distinction between criterion-referenced assessment (CRA) and norm-referenced assessment (NRA).

CRA focuses on the knowledge and skills a person has acquired - what the person can or cannot do - without reference to how he or she compares with other people. Criteria of attainment are usually described in terms of the content and objectives of the course. The most effective questions within a CR test are those which relate directly to the skill or concept being assessed. For example, in a driving test, the ability of a driver to perform a three-point turn is best measured by having the driver make a three-point turn in a range of standard driving conditions (e.g. on the flat, on a hill, with other traffic around, etc.).

NRA, on the other hand, focuses on the relative standing of candidates, that is, how well each person compares with others taking the same test or subject. Criteria of attainment are described in relative terms: for example, grades may be awarded such that candidates are grouped into the top 5% of pupils, the next 10%, the next 15%, and so on. Grades are often labelled as A, B, C, etc. The most effective questions in a NR test are those which heighten discrimination between individuals; such questions are not always very useful for telling us whether, or to what extent, an individual has mastered a particular skill.

In practice, it is rarely possible to assess student work without combining elements from both perspectives. For example, in order to set criteria to judge what students can or cannot do (CRA), an examiner usually draws upon experiences of what students typically or on average are capable of achieving (i.e. norms are invoked). Similarly, when NRA is used, tasks are set which usually specify criteria for judging the quality of students' answers in relation to their understanding of the subject; students are then ranked or grouped according to their position relative to other students taking the test.

Clearly, the issue is not black or white but rather which focus, CRA or NRA, should be given more importance. My contention, in line with *Tomorrow's Standards*, is that students have the right to be judged, as far as possible, on their own merits. If a student achieves a particular level on a task or test, that level should not be re-defined as higher or lower because the quality or level reached by enough other students is higher or lower.

However, I note a major inconsistency in the thinking of many people who support the emphasis on CRA. I include Dr Lockwood Smith who, in a recent "Perigo debate" on TV, supported CRA but also argued strongly for the retention of externally conducted examinations. He is not alone.

For the following discussion I will use School Certificate in its most common form, an externally conducted 3-hour examination in a range of different subjects, as my example.

School Certificate, while clearly containing elements of CRA, is predominantly NR in style. This is apparent in various ways. First, the examination prescriptions do not adequately specify performance standards. This is not to say that the examinations lack validity; simply, the prescriptions need to be in an entirely different form if they are to be used for CRA. Secondly, a greater proportion of the questions set within School Certificate examinations are of the sort which better provide discriminatory rather than criterion type information. Some questions, of course, are effective at both. Thirdly, the procedures used to scale marks focus clearly on the establishment of comparative norms rather than on standards defined in terms of achieved competencies. The scaling system in fact establishes a hierarchy of means between subjects based on the "general ability" of the intake of students in each subject. Thus a candidate's final mark in a subject reflects not only the candidate's own standing relative to those taking the subject, but also the standing of this group relative to the groups taking other subjects.

The question which arises from the preceding is whether the present School Certificate examination system can be modified or restructured to provide CR assessments of student performance. The central argument of the rest of this discussion is that CRA and externally conducted examinations are incompatible; CRA and internal assessment by teachers are not.

First, let us consider the option of running CRA by means of external examination. Under such a system, the initial task of the examining body would be to provide a detailed statement of the content and objectives of each course, probably running into several pages or even a booklet, and giving specific details of the knowledge, concepts, skills, applications and practical operations which need to be achieved. Depending on the subject and the viewpoint of the examiner(s), it is possible that levels of achievement would be specified with each level requiring attainment (therefore examination) before progress to the next. Ideally, the course prescription would be accompanied by a large sample of illustrative questions giving further guidance as to the nature and complexity of the performances required of candidates. Furthermore, the prescription would need to contain guidance concerning both its implementation by schools and the conditions (dates, regulations, etc.) relating to the administration of tests and other assessments.

If assessment is to be by means of external examination, the work of the examiner, the supporting committee and the team of markers would be substantially extended. In most subjects, I would envisage the need for at least three, and in some subjects as many as nine or ten, different tests annually in order to cover the course prescription in appropriate stages. This does not include the extra time and resources needed to construct alternative or "parallel" forms of each test so as to maintain security while providing a flexible testing programme; it should be borne in mind that pupils will not progress at a uniform rate. Also, since an essential feature of CRA is that student performance should be constantly monitored on an individual basis (teachers, not external examiners are in the best position to do this) rapid feedback on testing is essential in order to promote quick progress to later topics or levels. Consider for the moment the time taken for the Ministry to return School Certificate results (not scripts) to pupils and schools. Further, consider the effort needed to repeat this task, admittedly on a smaller scale (e.g. for a one-hour instead of a three-hour examination) three or even ten times a year. Now add to this effort the major administrative objective of returning scripts in the minimum time possible, ideally less than two weeks, if the system is to work effectively from the learner's point of view. I have probably said enough. By now you may well have jumped to the practical questions of employment and cost: more test constructors, more markers, more computer personnel, more administrators, etc. In short, a comprehensive system of CRA by external examination is impractical for reasons of scale, cost, effort and effectiveness.

As already indicated, it is my belief that CRA on a large scale can only be introduced within a system largely dependent on internal assessment. Teachers are in the best position to monitor achievement, to re-direct quickly student learning and to implement a testing programme at the appropriate steps in the learning process.

Unfortunately, any system of internal assessment has many practical problems to overcome notwithstanding the additional burden of operating CRA; for example, the task of monitoring standards is a complex one for which there is no one method universally suitable across subjects. It is in this context - the monitoring of national standards - that the role of the external examining authority is crucial. In short, under a system of CRA, the work of the external examining authority shifts from one of being the examiner of student performance to one of being the monitor of standards. It must put in place monitoring procedures that are relevant to each subject and which ensure that each school is treated fairly.

Accuracy of Educational Measurements and the Monitoring of National Standards

> *Assessment procedures need to apply consistent standards across the country.*

My conclusion at the end of the preceding section leads on nicely to the second theme of my paper, namely, that educational measurements are not of sufficient accuracy that we can make precise statements either about individual student performance or standards of attainment between schools. Whatever system of monitoring is used for a subject, the best we can hope for is "approximate" or "tolerable" comparability. The task is to achieve this with as little effort and cost as possible and not to become involved in a "law of diminishing returns" by trying to introduce complicated, time-consuming and expensive monitoring procedures which do not achieve significant improvements in our measurement of students.

In case you think that external examinations should not be discarded until we have ways of establishing more precisely the comparability of teachers' assessments, you may be interested to know that estimates of the "margin of error" for the marks awarded to candidates in external examinations are nothing to write home about: they typically exceed ± 5%.[1] These estimates of error are generally interpreted in line with the properties of the normal curve. This allows us to make probability statements about the "true" or "real" achievement level of a candidate when we know his or her actual score in an examination. For example, we should treat someone with a score of, say, 50 in School Certificate English, as having a "true" achievement level somewhere in the range 40 to 60. Furthermore, if we take this interpretation for all candidates who score 50, we would be right only 95% of the time; the true achievement level of the other 5% of candidates would be either above 60 or below 40![2] You might like to ponder this information in the light of the system used previously in School Certificate which passed a student who scored 50 and failed another who scored 48.

The evidence available on teachers' assessments indicates that their judgments are generally as reliable (I would argue more so in most cases) as those from external examinations, and are typically more valid because teachers are able to assess students on a much wider range of relevant skills and material than can be done in a three hour external examination. The main weakness of teachers' assessments is that interpretations of standards often differ between teachers (and therefore between schools); this results not so much in incorrect judgments about whether student A has done better than student B in a particular school, but rather whether the marks in general from one school are in line with those from other schools. Hence the need for monitoring procedures which are relatively effective and efficient.[3]

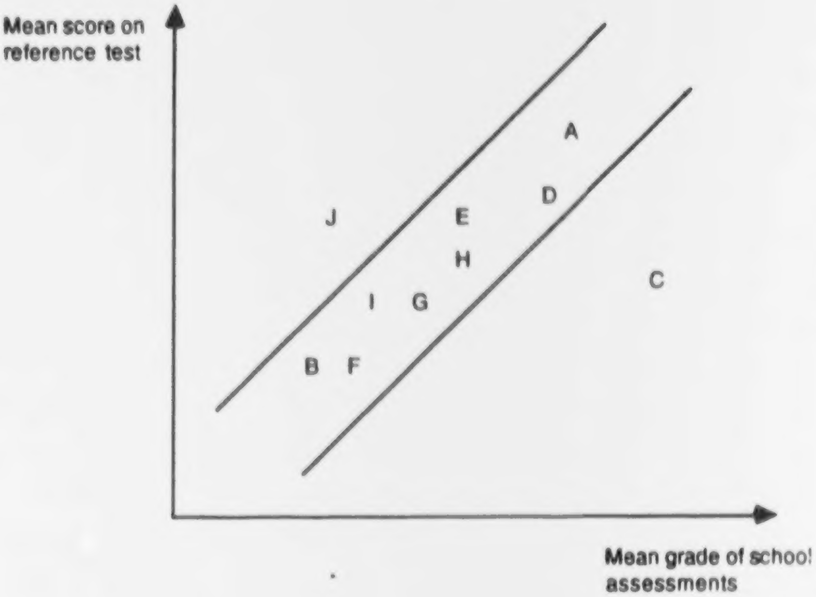
On this point I would like to turn to the suggested procedure for monitoring Sixth Form Certificate grades as set out in Appendix 4 of *Tomorrow's Standards*. The main features of the procedure are:

- A moderating test would be developed in each of the most popular subjects and administered to schools so that a pool of grades could be allocated to match the overall level and range of achievement in each school. Thus, a school whose pupils perform well on a test would receive a "better" pool of grades in a subject than a school whose pupils perform less well.
- The pool of grades given to a school would be assigned to individual students by teachers based on their own assessments of students during the year.
- The moderating test would be devised in such a way that it would provide continuity of content from one year to the next. This would provide a possible mechanism for monitoring shifts in standards over a period of time.
- The procedure would allow schools to alter their allocation of grades provided that the overall average for the school remained the same. In other words, the award of a higher grade for one pupil would have to be balanced by the award of a lower grade for another.

In terms of the criteria I suggested earlier, this model has positive elements - it is relatively cost effective and its procedure for achieving continuity over time is creative - however, the model is too deterministic in that it identifies what a school's grade allocation should be without checking the validity of the school's own assessments. In effect, what I am arguing, is that subject reference tests are a cost-effective way of monitoring standards, but they need to be used in conjunction with other procedures to ensure that each school is treated fairly.

Let's take a look at a typical comparability analysis (fictitious) using a subject reference test. The diagram below sets out the results of such an analysis.

Figure 1: Monitoring School Assessments Using a Subject Reference Test



The vertical axis identifies the mean score on a reference test obtained by the pupils from each school in the analysis. Schools are labelled as A, B, etc. In our example, schools A and D have done well on the test and schools B and F not so well.

The horizontal axis identifies the mean grade of the pupils in each school based on the internal assessments by teachers. Schools C, A and D have awarded the highest grades and schools B, J and F the lowest.

(Note that a more sophisticated model would also include comparisons between schools on the spread of the grades awarded, not just the average grade. For the present purposes, the above analysis is sufficient for explaining my objection to the model proposed in *Tomorrow's Standards*.)

Assuming that the reference test is a good monitoring device, a good correlation should exist between the mean test scores of schools and their mean grades. This is well demonstrated in Figure 1 by the diagonal representation of schools (from bottom left to top right). However, two of the schools, C and J, do not fit the pattern as well as the others.

The diagonal lines in Figure 1 identify the amount of "tolerance" needed to take account of the reliability of the educational measurements in the analysis (i.e. the reliability of the reference test scores and teachers' internal assessments).

Schools whose assessments fall within the tolerance lines can be assumed to be awarding grades more or less in line with the standards in other schools. Their assessments can be left alone. However, schools that fall outside this line need to be checked further. In Figure 1, school J appears to be somewhat "tough" on its pupils whereas school C seems to be "easy".

If reference tests could be assumed to be comprehensive in their sampling of the skills and content of the subject, the method of assigning marks described in *Tomorrow's Standards* would be sufficient to moderate school assessments. However, cases often arise where a school's assessments are comparable with the national norm but may not appear so from the performance of its pupils on a subject reference test. Such tests cannot assess the same range of skills that a teacher can do in school. Also, reference tests are not very sensitive to innovations by teachers which might, for example, promote independent thinking in pupils. Teachers also have available a wider range of assessment techniques, enabling them to better assess achievement of course objectives. And, in respect of schools with small enrolments, reference tests are not sufficiently reliable to give a stable picture of either the average or spread of achievement of the school's pupils.

In an analysis of the kind shown in Figure 1, further checks are needed before we can assume that the grades awarded in schools C and J need moderating. One method is to examine a sample of work from the schools to identify whether pupils are achieving at the level assessed by the school. While this would increase the time and effort in monitoring standards, it would likely provide a significant improvement on both the reliability and validity of the assessments of pupils. Samples of work would have to be chosen to ensure that they were representative of the range of work done in the school.

An argument might be made that all schools should submit samples of their work. As a system of monitoring standards, such a procedure would not be practical in the popular subjects because of the time, effort and cost in assessing such a huge sample of work. The method proposed here is that a reference test would be used as a first sweep to check standards; sampling of work would then be done for those schools which fell outside the tolerances set by the reference test.

Notes

[1] For example, research in Britain has shown that the margin of doubt, technically called the "standard of error of measurement" (SEM), typically exceeds ± 6% in GCE and CSE examinations (Willmott, A.S. and Hall, C.G.W. (1975) *O Level Examined: the Effect of Question Choice*. Schools Council Research Studies. Macmillan Education; Willmott, A.S. and Nuttall, D.L. (1975) *The Reliability of Examinations at 16$^+$*. Schools Council Research Studies. Macmillan Education). Similarly, evidence I pieced together in the mid-1980s (unpublished), suggests that the SEM for School Certificate subjects probably lies somewhere between 4 and 7 marks (4 in the case of Mathematics and 7 for Woodwork and other practical subjects).

[2] These figures represent the 95% confidence interval for a particular score in an examination. The 95% confidence internal is determined by: Obtained score ± 2 x SEM.

[3] Evidence about the level of agreement between teachers and external examiners is given by, for example, McCausland, F.J. and Hall, C.G.W. (1985) "The Accuracy of Teachers' Predictions of Pupils' Marks in School Certificate". New Zealand Journal of Educational Studies, 20, 1, pp. 82-92.

Marilyn Davies
Education Adviser
New Zealand Employers Federation

I have been asked to represent the views of employers on national education monitoring in general and on proposals contained in *Tomorrow's Standards* in particular.

My focus will be on those parts of the report relating to the senior secondary school and dealing chiefly with School Certificate, Sixth Form Certificate and Records of Achievement moderation.

I have had a number of meetings with employers throughout the country on achievement-based assessment, on the ABLE discussion document and on the Record of Achievement project. Certain recurring themes come through discussions. The fact that these ideas are heard from companies large and small, over all industries and sectors, in all parts of the country, and are supported by overseas experience and literature, tends to suggest that the messages are not simply ritualistic put downs of the education system. These are genuinely concerned people who, given that education is a service sector industry, happen to be our clients and customers.

I'd like to look at these recurring themes and view *Tomorrow's Standards* in light of them.

1. There is a broad range of basic or core skills which employers expect young people to have acquired by the time they leave school.

   In general terms these cover such areas as communication, information handling, problem solving, number, technology and computer skills, the physical, practical and creative areas, and a range of work and social skills such as teamwork, reliability, self-organisation, initiative, perseverance. Some would call them the new basics or the 4 C's: communication, computation, critical thinking and computer literacy.

What is taught is less important than how it is taught and learned, how it is measured in terms of achievement levels and competence, and how it is reported. Employers are interested in the knowledge base and content of what is taught. Indeed they need to be assured that all students command enough shared knowledge as a basis for further learning and to be able to communicate effectively with others. But the terms of reference of the ABLE working party did not extend that far. Discussion on assessment and reporting is therefore being carried on in a vacuum. We are looking at the outcomes, rather than the processes of the school system. In these terms the Record of Achievement as currently being trialled in nine state secondary schools goes quite some way to fulfilling employers' need to have better information about the achievements and capabilities of young people.

I have three comments to make on the Record of Achievement proposals.

(i) It seems illogical to sever the link between a record of achievement and the recording of national awards.

We have been told it is impracticable, partly because the documents may be needed at different times. I don't buy this argument entirely. In this day of rapid turn around of information and delivery of services, the hiatus between information wanted and information supplied is becoming shorter. In any case it's not much different from the sort of situation which currently applies with July or August applications for some tertiary courses.

The inclusion of a summary of national awards acts as a moderating device and supports the credibility of the school-based profile. This brings me to my second comment.

(ii) The question of moderation - a tough one, but crucial for the customers. It's a quality assurance process which helps to make the document believable.

I appreciate the efforts made by the working party to institute moderation procedures for the Record of Achievement. The difficult nature of the task is reflected in the proposals relating to national testing of basic skills at fourth form level.

While these may be desirable and appealing on paper, and deserve your serious and close attention, the questions of cost, and educational effectiveness may militate against their feasibility. In addition, the use of fourth form skills tests to moderate a leaver's record a year or two later smacks of the present system of School Certificate moderation of Sixth Form Certificate. I should nevertheless like the feasibility of common skills testing to be investigated for moderation purposes. It may be, however, that the question of broader-based moderation for a record of achievement has to revolve around:

a) standardisation of language and format;

b) nationally agreed interpretation of descriptive skill levels, supported by exemplars, other resource material, in-service courses, and checks by Education Review Officers;

c) inter-teacher moderation within the school;

d) as mentioned earlier, the inclusion of a summary of national awards, or credits towards them.

This brings me to my third and final comment about Records of Achievement.

(iii) I want to move from the particular to the general and paint a broader picture of the Record of Achievement - one in which the Record is a cumulative document for life, incorporating discrete items such as Primary Progress Records, the school leaver profile (i.e. the present Record of Achievement), the proposed National Certificate of Education, and tertiary awards.

In principle, a cumulative Record of Achievement of this nature would be owned by the individual and carried with them to assist transfer from one learning stage to the next. It would be complementary to a national qualifications framework and would support the notion of credit accumulation and transfer - but I am stepping outside of my frame of reference. That is for another forum.

2. A second recurring theme from discussions with employers is the need to upgrade nationally recognised standards of achievement.

New Zealand is ill-prepared to face the demands of an internationally competitive economy, in which the demand is for goods and services which feature high quality, rapid delivery, uniqueness and cost effectiveness. To deliver the goods in the international environment New Zealand needs a more highly skilled, highly educated and highly qualified population. I am not here to get into debates on so called 'official' and 'radical' views on the inter-relation between jobs, skills and the economy. I simply want to try to unpack, from an employer's point of view, the concepts of 'standards' and 'achievement'. Employers adhere strongly to the notion of a nationally coherent education system.

With increased retention rates into year four, more interest is developing in Sixth Form Certificate and achievement-based assessment. Although there is considerable confusion about what norm-referenced examinations measure, more and more employers do understand that norm-referenced grades do not measure standards or tell them anything about levels of knowledge and skills, or the criteria against which the student was assessed. Assessment based on descriptive graded criteria will do much to appease increasing concern over the issue of standards, since the criteria define standards at each grade and the student is awarded the grade that matches the required level of performance. But the customers are inclined to see anything that is not externally examined as lacking in rigour. We need to market systematically this type of assessment.

The marketing should be based on concepts of 'standards' and 'competence' - concepts familiar to employers, although misunderstood by many. The understanding will be enhanced, however, if quality control processes are seen as integral to the assessment system. I'm talking about the concern for national moderation which centres upon:

- the need to ensure that the criteria are being interpreted uniformly and consistently
- the need to ensure inter-school comparability for reasons of fairness for students as well as for end users of certification.

I'm not going to argue now for one particular type of moderation procedure over another. I have heard arguments for and against the Appendix 4 proposals in *Tomorrow's Standards* - I leave it to the experts to sort out the most effective mode, whether reference testing, visitation or some other type. My message is that moderation must foster public confidence in a nationally based certificate. It is seen by employers as an independent check on the maintenance of national standards and is essential if achievement-based assessment is to work in New Zealand.

No discussion on standards and assessment would be complete without reference to School Certificate. The Employers Federation issued in September 1989 a policy statement on School Certificate. It stated, essentially, that before School Certificate is removed other measures need to be put in place to assure employers and others that the achievements and abilities of school leavers are effectively measured and documented. I have to state that a majority of employers surveyed by us continue to favour an external national examination or standard test procedure for all schools at fifth form level, or at the completion of compulsory schooling. There is still a strong correlation in the minds of many between external assessment and standards, and a view that only external assessment is fair and unbiased.

There is evidently a fair amount of promoting and marketing to do, but I do not believe it should include negative marketing of School Certificate. A two year campaign for public recognition and acceptance of Records of Achievement and Sixth Form Certificate achievement-based assessment and moderation, combined with increased retention rates, is far more likely to result in a lowering of the market value of School Certificate.

Perspectives on National Monitoring

David Greig
Deputy Manager
Social Policy Branch
Treasury

Many of you today are discussing the issues raised in *Tomorrow's Standards*. In the Treasury we are still thinking through that report, so what I can say today has a personal rather than official status.

I will start by outlining the Treasury's role, both generally and in education. Then I will discuss the general thrust towards monitoring of performance, not just in education but in the wider public sector. Finally I will comment in more detail on monitoring in the education sector and make some observations on the current debate.

Much of the Treasury's role is defined by the new Public Finance Act which has changed the game a lot of us are playing. Previously, both in the Treasury and throughout the public sector, discussion about the use of resources, including in the education sector, was based on inputs such as staff and buildings. For example, the debate about pupil/teacher ratios has a clear focus on inputs.

The Public Finance Act obliges us now to base our assessments on outputs instead - that is the level of goods and services produced. In the case of education, outputs in a simplified sense are the number of students multiplied by the number of years in the system - and this can be broken down into sub-categories. We focus on three aspects of outputs: their quantity, their quality and their cost.

Another new element in the Public Finance Act is outcomes, which are usually Government-determined views on what is to be achieved by the outputs. In education, outcomes refer to skills or other aspects of learning.

A twist to this in the education sector is that it is very difficult, except in the classroom itself, to make direct assessment of the quality of the output - the output being the child being taught. We have to use a proxy for that, which is achieved by looking at the outcomes. An assessment is made of the student's performance and from that a judgement has to be made about the output. The judgement can become complicated, depending on the other factors influencing the results.

At the Treasury one of our main roles is advising the Government about the quality, quantity and cost of outputs - or to put it in plainer language, value for money - and of course we do that for all sectors of the economy; the Education Section in the Treasury has only 3 staff out of over 100.

The analytical task that we and other people undertake is to try to get a view on the relationship between the outputs and outcomes. Another major role which we have of course is the more directly financial role - the analysis of the total cost and what that means for the economy, but that is not what I am here to discuss today.

In education the easier aspects of the outputs/outcome task concern the quantity and the cost of the outputs. There are plenty of small scale problems in assessing those, but basically we and the others we deal with know fairly well what outputs are costing and how many students there are at the different levels, apart from one or two difficulties up in the far north.

The more difficult task is quality, which is what this conference is about. It has been especially difficult in the past because, as *Tomorrow's Standards* says, in a strict sense there has been no national monitoring at all - national in the sense of being completely external and nation-wide. We all have to work with the results of internal assessment or assessments which have the norm problem others have been talking about this morning. This really is a problem for anybody outside the education system.

Our view, and this is not for education alone, is that we cannot simply rely on assurances made by providers irrespective of whether the activity is manufacturing, banking, education or whatever. This applies even when there is no doubt whatsoever about the integrity and professional competence of whoever is giving the assurances. It is hard to have confidence when one is entirely dependent on assurances from within an industry or a sector. For example, we were never at ease in the debates, especially in the 1960s, in which manufacturers assured us that they were doing a good job and therefore imports should be restricted. We were not at ease about suggestions from banking circles that they were efficient, that their margins were low and therefore there was no case for allowing new banks to set up. The same applies here. We have to be uneasy about depending too much on internal monitoring of education.

We have the same sort of problem even when thinking about ourselves. How can anyone figure out whether the Treasury is doing a good job? It is not something you can measure very easily. We have internal mechanisms such as hierarchical controls and internal peer review, but we also often have the important element of external peer review, by qualified people outside the Treasury.

The pressure for greater attention to national monitoring of education performance is widespread in Western countries. It has been particularly strong in the United Kingdom and the United States; New Zealand is not alone. I also think education is not alone. The push is wider across all of the public sector and it is not an aberration. I find it interesting that the push is coming from outside the education sector. It is coming from parents and Governments and, as we heard earlier, from employers - generally people concerned about the quality of the educational outputs and whether the education sector is doing a good job for the country. In some countries, notably the

United States, there is concern about whether the quality of education is declining, and in many countries including New Zealand, there is interest in how well we are competing in an educational sense, with other countries. There is also increasing awareness of the problem of children who drop out of the system and end their schooling unemployable.

In my view the pressure that we are all facing for increased accountability for our performance reflects some fundamental changes. The main Western countries are facing tougher competition than they did. The range of unskilled jobs available has shrunk. There is a demand for skills to help the economy not only to grow, but to compete with others who have overtaken us.

I do not claim any expertise in the technical aspects of the assessment of education outputs and outcomes, but would like to offer a few observations about the current debates.

The first is that whatever assessment systems are in place have to be credible to people outside the education sector, not just to those inside it. The obvious outsiders are parents, employers and the Government, which is the provider and funder of most of our education. I have not followed the debate as well as some of you, but it seems to me that acceptable credibility does require a degree of external national monitoring. The same can be said for other activities outside education. The monitoring must have a substantial and credible external, or neutral element in it. If the system is not credible, the reality is that others will do it instead. We are already seeing that, with the Auckland-based Education Foundation developing exams which various schools are using. It is not realistic to try to stop that sort of thing happening as people will ultimately exercise a choice.

The second observation, which picks up a point made earlier this morning, concerns the argument about assessment causing narrowness in teaching and learning, producing the opposite result from what was intended - a decline in performance, not an improvement. The argument is especially prevalent in the United States. I think that it is a real argument, but I do not see it as an argument against external assessments as such. I see it as an argument for being very careful about how you monitor, and exercising care in the design of whatever assessment procedures you have so as not to cause that narrowing effect. We have heard the same sort of idea from the panel just now.

The third rather obvious observation is cost. The U.K. is now finding, having decided that they would like to have tests at seven, eleven, fourteen and sixteen years, that they are not only costing a lot of money, but are cumbersome to put in place. I didn't detect any discussion about cost in my reading of *Tomorrow's Standards*, but it is a concern when thinking about nation-wide testing of all children at certain ages.

The next observation I have is I think the most controversial. It concerns the debate about failure, which we saw a few years ago with the old School Certificate arrangement whereby half the students failed and, therefore, half became demoralised. I accept that argument as far as it goes. But the other extreme which I've heard sometimes, that all assessments should be positive in some sense, strikes me as really going too far. The hard reality is that society and the economy have to sort people out, so they do. I agree they should not just sort them out in some black and white sense, such as academic/non-academic. People have a vast variety of different abilities. It is a multidimensional question rather than a linear one. If you look at it in another way, everybody at this conference is successful in some activity related to education, but I guess everybody in this room could also be seen as a failed nuclear physicist or a failed bulldozer driver. It's not a negative thing to sort out the roles that people can play most usefully and where they will get the most satisfaction. If schools don't help society with this sorting then others outside will do it, and maybe not do it very well.

I see the so-called failure question really as one of providing courses that different types of students will find challenging and interesting. In the old days we had a small proportion of students going on beyond about the age of sixteen and that small proportion were mainly of an academic orientation. Now there's a much higher percentage of students going on, and the system is responding by developing different types of courses, which can be assessed. There's also a question which I didn't see addressed in *Tomorrow's Standards*, which is the one of providing interesting challenges to exceptionally bright children.

Another observation, on which we've heard some discussion earlier this morning, is the question of unfair interpretation of the results. This arises especially with inter-school comparisons, and I agree that if you have one school which has a struggle because it has to deal with a lot of children who have had very difficult family backgrounds, and another school which has an easy time because of the material that's fed into it, then the straight comparison would give a misleading impression of how well each of those schools did. But the solution to that is not to stop making these comparisons, but to see to it that results are presented along with an analysis and a discussion of all relevant external factors. Comparisons are going to be made anyway, and everyone has had the experience of how people in the community compare different high schools - often with a low standard of discussion from which strident conclusions are reached.

One final observation is to be beware of overdoing national monitoring. There's a danger of a grading and monitoring industry, which uses hundreds of indicators, places high costs on everybody involved, and does not know what it is really for. The gathering of indicators nation-wide is a very expensive and disruptive process and should only be done with a lot of forethought. The aspects I am particularly sceptical about are inputs-related indicators - how many square meters per child, how much do you spend on different types of activity in the school? People will generate hundreds of indicators which they think might be interesting because they might give you a clue about the effectiveness and the performance of the school. However the relationship between those inputs - how you use your money for equipment, staff, buildings etc. - and the output at the end, is complex and there are hosts of different views about it. If you prescribe the set of indicators that you gather in, you're almost prejudging the relationship. I think a far healthier approach is to be very restrained, gather national information only on basic indicators, and concentrate on sampling and on ad hoc gathering of information, for specific research purposes.

National Monitoring in a Treaty of Waitangi Framework

Katarina Mataira
Waikato Polytechnic
Maori Language Commission

What a frustrating word is 'tomorrow' for policy makers, education administrators, teachers and parents, presently in the throes of restructuring education. The dawn of the morrow is expected to herald a new enlightenment - better education for all, and a brighter future for the nation. Meanwhile, headaches abound and some people are wishing that 'yesterday' might come back.

With the release of *Tomorrow's Standards*, we are allowed a peek at the machinery which will very likely be put in place to make absolutely certain, that the promise of 'Tomorrow's Schools' and our expectations for the brightest of tomorrows will be fulfilled. The Minister of Education obviously approves and the tax payer can expect to pay two and a half million dollars a year for the privilege.

How then do I, an ordinary teacher, dedicated to the value and the purposes of education in its widest sense, respond to the recommendations of this report? Further, how do these recommendations relate to the aspirations and expectations of the Maori people in a Treaty of Waitangi context?

There can be no question that the assessment of students is long overdue for an overhaul. Certainly, the recommended shift away from assessing students against their peers, to assessing them against agreed criteria and clearly defined levels of performance, ought to signal more clearly whether or not educational progress is being made. There is still, however, the question of national criteria and national standards against which individual students will be measured. How will these criteria be determined and who will set the standards? If the Ministry of Education is going to be responsible for national monitoring, as recommended by the report, will they be setting the criteria and the standards for assessment? And is there not a danger that the final outcome will be something similar to norm-referenced criteria?

As for assessing the effectiveness of schools and teachers - it is not surprising that the report recommends that this task be undertaken by some other Ministerial working party. No school or teacher will welcome yet another massive change to their lives. At the same time, it is abundantly clear that both these issues have to be addressed. The report at least recognises the need for changes in the training and development of teachers.

What about Maori students, Maori teachers and Maori schools? In particular, what about Kura Kaupapa Maori? The report is actually very circumspect in its recommendations with respect to 'assessment and New Zealand's dual heritage'. What is to be gained by an investigation into the ways Maori children answer the items in a standardised test? Or the setting up of a body to oversee educational assessment as this affects the Maori school population? Or an agency to report to the Iwi Maori on issues pertinent to assessment?

If the function of any such body is to be a mouth-piece for the Ministry and its policies on educational assessment, how will Maori parents respond? Their disenchantment with mainstream education is well documented, and brown-skinned persons preaching yet another Pakeha scheme for the salvation of Maoridom will serve no purpose other than reinforce the alienation of Maori people. In the context of the Treaty of Waitangi, sovereignty is paramount.

With respect to the education of their children, Maori parents have divided into three separate camps - those who are happy to have their children in mainstream schools, those who are opting for bilingual schools, and those who are committed to Kura Kaupapa Maori.

Kura Kaupapa Maori are presently few in number but it is my contention that their numbers will increase as Maori parents assert the right of their children to be educated through the medium of the Maori language. In these schools there is no question of any child being neglected in any way.

Will Kura Kaupapa Maori raise the levels of achievement of Maori children? They are absolutely committed to that task. And given the resources, I believe they will do just that. They have but one overriding concern. The education authority which denied Maori children their language 'yesterday' may well deny Kura Kaupapa Maori the right to pursue 'tomorrow' what they implicitly believe is the best possible education for their children.

The *Tomorrow's Standards* report recommends that decisions about Maori education must come from Maori people. True - as long as the Maori people whom the Ministry consults are not only those within their own camp, or those committed to mainstream education. Kura Kaupapa Maori are a special case and ought to be heard in their own right. That right must include decisions about performance criteria, levels of achievement and assessment procedures appropriate to the needs of their children, the expectations of their parents and the goals and objectives of their school.

The principles of sovereignty, equity and partnership encapsulated in the Treaty of Waitangi, can be realised in Kura Kaupapa Maori. With respect to assessment procedures for Kura Kaupapa Maori, sovereignty is acknowledged when they are accorded the right to determine the criteria and the standards against which they will be assessed.

Equity is realised in their receiving a share of the resources for the establishment of an appropriate culturally sensitive assessment system. Through making both sovereignty and equity a reality in practical terms, power is shared and partnership achieved.

The Opposition's Response to *Tomorrow's Standards*: Policies and Initiatives

Dr Lockwood Smith M.P.
Opposition Spokesperson on Education*

In my kinder moments, rather than describing *Tomorrow's Standards* as a bit of a dog's breakfast, I would say it is very much the work of a committee - something for everyone. I know that statement won't have endeared me to many of you, but assessment in education is too important to pussyfoot around. I look forward to healthy debate on what I've got to say to you.

The fact that *Tomorrow's Standards* is unfocused, is not entirely the fault of the working party. The Terms of Reference provided by the Government were flawed. Like much of the *Tomorrow's Schools* upheaval, the Terms of Reference failed to put the pupil first.

The first requirement of the working party was to "recommend to the Government procedures for assessment which:

1. Can monitor the effectiveness of the New Zealand school system on student learning;
2. Assess the effect of individual schools on students' learning achievements;"

They also had to recommend ways of reporting those assessments and within the context of New Zealand's dual cultural heritage, advise on the possible effects of such assessment. The emphasis was on the school system and on schools, with a vague requirement to outline the main issues of assessment for better learning and to describe a range of models and procedures for assessment. There was no specific mention of the need for assessment proposals to better determine learning need and to identify children requiring additional teaching help. That is why I charge the Government with failing to put the needs of pupils first, despite the rather grand name of the working party.

---

* Since delivering this paper Dr Lockwood Smith has been appointed Minister of Education.

The working party interpreted these terms of reference saying, "we were required to make recommendations on assessment procedures for evaluating and monitoring the education system as a whole, as well as for schools, teachers, and students." And so before the working party got into action, the vague terms of reference meant that assessment of schools and teachers was to become confused with the monitoring of students' learning. Little wonder the final report lacked focus.

The Labour Government established the working group in April 1989. The Government didn't have much choice. For 12 months, since the Picot Report had come out, I had been arguing in speeches up and down the country of the need for new assessment systems to enable us to monitor pupil achievement nationwide. It was something the Picot Taskforce had largely ignored. Parents concerned about the lack of information on their own children's progress at school supported my calls. There was increasing public dismay at a perceived lack of achievement of competence in literacy and numeracy in our schools. In December 1987 the publication of the Association of University Teachers, *The Bulletin*, carried, under the heading "Falling Standards of Literacy/Numeracy", the following statement.

> "Earlier in the year, the Auckland Law School reported that it had become necessary to introduce remedial English classes for the vast majority of new entrants, as it was felt that law graduates would soon not have sufficient command of the English language to be competent practitioners. Similarly, Science Deans reported the need to establish bridging courses to enable 7th formers to cope with Stage 1 classes."

What's frightening about this, is university students are supposed to have been the successful ones in our education system. That same year, the Department of Education surveyed Access trainees. They found that 60% or 70% of them were neither functionally literate nor numerate. In the face of this superficial evidence of a major literacy problem in our education system, educationalists cite more formal studies, such as the surveys of the International Association for the Evaluation of Educational Achievement which show standards of literacy in New Zealand to compare most favourably with other countries in the western world. Pity about mathematics.

But, of course, that begs the question - what has happened to standards of literacy in the western world in the past two decades. There is concern that those standards have declined. Even *The Economist* reported in January this year that at Motorola, an American electronics firm, 4 out of 5 applicants for jobs fail an entry level employment test that requires the English competence of a 12-year-old and the Arithmetic competence of one still younger. *The Economist* reported that McDonalds has had to put symbols of hamburgers, cokes and fries on its cash registers because many of its fast food servers cannot cope with dollars and cents.

If you doubt the public concern, just talk to secondary school principals about the literacy and numeracy competence of their third form intake over the years; talk to the Polytech tutors about the basic educational attainment of young people entering our Polytechs. They tell me they may have to impose entrance exams because they can no longer rely on school qualifications. Like it or not, there is widespread public concern that standards have declined. Sadly, if it's not true, with present assessment systems we don't have the evidence to properly refute it. In this regard, I have no argument with the proposal for monitoring national performance.

A sampling of educational outcomes at Standard 2 and Form 2, at intervals of 4 years, would give us more information on any changes in educational standards in our primary schools over time. The proposal that all fourth form students should each year be assessed on important basic skills, however, makes no sense to me at all.

The working group's second set of proposals on assessing the effectiveness of schools, is woefully inadequate. This is a complex area. Given the recent work of Lauder and Hughes, and Nash and Harker here in New Zealand, and Smith and Tomlinson in Britain, it is quite clear that there is a significant school effect, but the complexity of the factors that interact to produce this requires sophisticated design and analysis techniques to unravel them. The simplistic education performance indicators proposed in the section on assessing the effectiveness of schools are quite inadequate. But I'll come back to that later. It is on the assessment of the attainment of students in schools that I want to focus most particularly this afternoon.

The working party has quite understandably separated out primary schools from secondary schools. For primary schools they propose that teachers should maintain a folio of each pupil's work, that each pupil's progress in reading should be carefully examined at age 6, and that the educational and social development of each child be thoroughly reviewed by teachers and family members at about the time the child turns 7, 9 and 12.

Against which criteria is it proposed the education and social development of each child be reviewed? How are parents meant to understand whether or not their child is making satisfactory progress? What is more, if the child is not making satisfactory progress, nothing has been proposed to remedy the situation. Surely assessment, that ascertains whether or not a child is progressing adequately, must be linked into appropriate responses so that the child can be helped if it is not making that progress. Therein lies my greatest criticism of the proposals for the assessment of students in primary schools.

During those early years the fundamental reason for assessment should be to determine which pupils need extra help and to incorporate into the overall assessment process, a mechanism for delivering that help. National argues that two groups of children need additional help. They include the pupil that is falling behind and not making the progress they ought to, and the other group is gifted children. Too often they don't get the extra help they need. Consequently, assessment procedures should tell us where in the school system those groups of children are so that we can deliver more resources for the schools to provide the help they need. That help may take different forms. It may be smaller classes. It may be an extra Reading or Mathematics Recovery teacher.

It's on this basis that we've developed National's policy we are calling the *Achievement Initiative.*

That policy has three clear principles. First, in the area of English, Mathematics and Science/Technology, we will define the learning goals, or attainment targets, that pupils should normally be seeking to achieve right the way through the core school system from 5 to 15. At present we have no such continuum of learning goals and consequently the school curriculum and syllabuses don't integrate properly.

The second element in the achievement initiative is new nationwide monitoring procedures to monitor pupil progress against those learning goals. We want those monitoring procedures to be as consistent with normal classroom assessment practice as possible. We don't want young primary school children sitting down to examinations. That is an impression my Labour opponents have been trying to create.

The third and most important element in the policy is additional resources to meet the learning needs shown up through these new monitoring procedures. In other words, special needs funding for schools to help schools address the learning needs of their pupils will no longer be distributed on the basis of race or social class, but on the basis of reasonably objectively assessed learning need. This is important because, after all, learning need knows no racial boundaries. Harker has shown also that learning need can vary substantially within a given class in society.

At present, schools like my Ruawai Primary School, might have only 30% to 35% Maori pupils on their roll, and many of the parents involved in farming, and get not one dollar of special needs funding, yet have desperate learning need that they struggle to address.

Where this proposal differs substantially from *Tomorrow's Standards* is that parents will know how their children are progressing against clear learning goals. The checkpoints won't be at 7, 9 and 12, but rather, at points where it is most appropriate to provide that extra help to a child to enable it to catch up before it becomes swamped by a feeling of failure. In National's achievement initiative there will be a general evaluation of all children entering school at age 5. At age 6 we have the Reading Diagnostic Survey and that will be continued. English will again be monitored at age 8. The checkpoints for Mathematics and Science/Technology are yet to be determined but will relate to the most appropriate time to provide recovery action.

There will again be a general monitoring at age 11 on entry into intermediate schools and at age 13 on entry into secondary school. What on earth is the point of having the review of a pupil's progress at age 12, as recommended by *Tomorrow's Standards*, when by then it's rather too late for the primary school to do a great deal about it. It's the secondary school that will be involved in the further needs of the pupil and the monitoring should take place on entry into that secondary school. The recommendations of assessment of students in secondary schools will not be pursued by a National Government, although I promise you I will listen if you have compelling arguments in favour of the proposals. As I've already mentioned, I see no logic in the national basic skills testing at fourth form level. It's too late to help the students involved, it doesn't provide them with any qualification and in my view is wasting resources. Again the proposals reflect muddled thinking.

There is no way that one certificate can meet the qualification requirements of all senior secondary school students. Those going on to university require the Bursary and Scholarship examinations. The concept of a cumulative National Certificate makes good sense for students going on to vocational education and training, so that they can build on that National Certificate at polytechnic level. It is National's clear policy to bring in such a National Certificate and once that is in place, Sixth Form Certificate will reduce in importance, simply being a period of study on the way to Bursary and Scholarship. National also proposes a new *Training Opportunities Programme* that will supercede Access, and be available to senior secondary pupils who have got through to the sixth and seventh form level with minimal skills.

A National Government will, however, retain School Certificate. The built-in failure inherent in the scaling procedures will go. The standards required for School Certificate will be published and the examination will be set each year based on those standards. Considerable effort will go into moderation to ensure that the examination does not vary from year to year. Students will receive their actual marks so that the efforts of both teachers and students can be properly rewarded. The maintenance of School Certificate and the Bursary examinations is most important to ensure that students are not disadvantaged by the school they attend. A pupil obtaining good marks in School Certificate from my local Ruawai College has got just as good a qualification as a pupil with the same School Certificate from Auckland Grammar. If that examination is done away with, and qualifications in the senior secondary years become totally internally assessed, my students at Ruawai College will be severely disadvantaged. The school a student attends will start to take on much greater significance.

It should be apparent from what I've shared with you that National views assessment quite differently at primary and secondary school level. At primary school level the main purpose of assessment should be to ascertain how pupils are progressing against learning goals so that additional help can be provided where it is needed. At secondary level, the main purpose should be to ascertain whether or not a student has met the requirements to receive various qualifications. Except for modules in the Training Opportunities Programme and National Certificate, the greater portion of that assessment under a National Government, will remain through external examination.

At the recent second New Zealand Conference on Research into Educational Policy, there was some discussion over that policy stance. When I put it to the conference that the public and particularly employers viewed examinations as the fairest way of determining attainment in the senior secondary years, some educationalists suggested that employers should be educated until they accepted internal assessment. I hope the suggestion was made in jest. The international marketplace and the world of work knows very clearly what it requires. We educationalists should listen and learn, or else we do young people a disservice. It is their standard of living in the future that we risk jeopardising if we ignore what that marketplace is telling us. At that conference, a member of the council of the School Trustees Association pointed out that if external examinations are dropped by the state, private organisations will reinstate them. Each year we see more and more schools joining the sixth form examination, and separate scholarship examinations have now been offered by the Education Foundation, a private organisation funded by business, and a surprising number of pupils have registered to sit the examinations.

I finally want to return to the complex issue of assessing the effectiveness of schools. It is in this area that I believe the working group's recommendations were at their weakest. They essentially conclude that "no value added methodology in existence now, or in the foreseeable future can be used for routine reporting or for ranking of schools." They go on to recommend that the Education Review Office should monitor overseas and local investigations into ways of measuring the effectiveness of schools.

There is increasing evidence that this is an inadequate recommendation. The work of Lauder and Hughes here in New Zealand and Smith and Tomlinson in Britain, shows massive differences between the attainments of students in different secondary schools, even when those attainments are corrected for intake characteristics, such as social class and measured ability through regression analysis. The New Zealand and British work appears to be at odds in terms of the impact of social mix where the New Zealand work shows mean attainment in schools to be depressed where there is a high ratio of lower socio-economic background pupils to high S.E.S. pupils. In apparent contrast, the British research reported pupil attainment did not depend on the numbers of ethnic minority pupils attending, ethnic minority pupils tending to be from lower S.E.S. backgrounds.

The work of Nash and Harker, at Massey University, demonstrates that yet another factor may be at work. They argue that the literate culture of families may have a more significant impact than their actual socio-economic status.

In reviewing the British research of Smith and Tomlinson, Dr Nash recently wrote that "differences in progress of the order of a standard deviation between pupils of similar ability, after only two years, due to the school they attend have never been reported before, and seem to challenge the conventional wisdom that schools are relatively powerless to interrupt class-determined differences in attainment".

From this research work it is quite clear that schools can have a profound effect on the attainment of pupils. It is not good enough to say that the Education Review Office should monitor overseas and local investigations into ways of measuring the effectiveness of schools in the way the working party has proposed. The possible impact of having more schools operating as well as the most effective schools is too profound to leave this further study up to the Education Review Office. The next National Government will further investigate these issues. Through our achievement initiative there will be more comprehensive data available on the attainments of pupils entering the different segments of our school system. The opportunity for comprehensive study of the school effect will be enhanced.

A National Government will want to see, and will be prepared to back, quality research into the effectiveness of early childhood education in our schools. The universities will be involved in research associated with our *Parents as First Teachers* development, that will work on literate culture with families, and both the universities and the Council of Educational Research will be most welcome to be involved in further research into the school effect.

We owe it to our young people not to put these difficult problems in the too hard basket. Unravelling complex processes is always difficult, but because of my own scientific background, and the way the mysteries of nature have been unravelled, I refuse to accept that we cannot better identify the factors that can help all pupils achieve more.

There is much to be done. I look forward to working with you to see it done.

Symposium on Opposition Spokesperson's Proposals

The task of each speaker in this second symposium was similar to the one asked of each speaker in the earlier symposium, namely to comment on *Tomorrow's Standards* and discuss particular points from the Opposition Spokesperson's address. This aspect of each contribution was always going to be demanding too, as Dr Lockwood Smith's paper was not available prior to the conference.

Each respondent has made important points in reply and has identified different aspects of *Tomorrow's Standards*. Shona Hearn sketches in the political background to *Tomorrow's Standards* and introduces the value-added notion discussed in the report, including the need to develop an 'acute methodology'. She notes that implementing national monitoring at Form Four may well see four consecutive years of secondary education dominated by formal assessment. She points also to a series of initiatives undertaken by previous Ministers without discernible outcomes, and laments the loss of valuable 'institutional memory', as a result of 1989 restructuring. Lockwood Smith's proposals for senior secondary school examining are discussed and three areas for investment in education are identified as needing to take priority over national monitoring.

Ross St. George reviews other major reports and working parties in recent years and asks whether *Tomorrow's Standards* will be shelved like some of these. He raises five themes from the report for further discussion; the time-lag associated with information coming from national monitoring, the need to have assessment procedures open to public scrutiny, the importance of investing in improved training in assessment, a tendency to label the achievement of sub-groups within the school system as a function of the assessment process and the need to clarify the key terminology of assessment. The questions raised by Ross St. George are timely, and make an important contribution to the scrutiny *Tomorrow's Standards* should receive.

Ken Wilson has made his response on behalf of the Trade Union Education Authority and this brings a union perspective to the ABLE report. His opening point is that New Zealand is a classed society and that those who identify themselves as working class may not see schools as necessarily representing their culture, interests, or priorities. He suggests too, that one of the influences underlying moves to national assessment may be a desire to divert attention from youth unemployment to education, and promote this debate by scrutinizing the performance of teachers and schools. He concludes that Project ABLE proposals offer some opportunities and less disadvantages to a wide spectrum of youth. This particular broader social view provides a valuable complement to much of the educational debate in these Proceedings.

Shona Hearn
President
PPTA

I want to begin with a few remarks on our perspective on ABLE before I move on to talking about what Lockwood has said.

I wasn't a member of the ABLE Committee but I have been a member of similar kinds of committees and I know that it's sometimes easy to forget that the whole thing has been an immensely political process. It starts, as Lockwood Smith indicated, way back in the ideological battle over Terms of Reference and who actually gets onto the Committee in the first place and it continues as a process of negotiation throughout.

It seems to me that considering those factors and the kinds of pressures that were on the Committee, that the document has a surprising amount of integrity and contrary to what Lockwood Smith said, I would suggest that there is a fair amount of concern for the needs of the students evident in the document.

I note that they're reasonably honest about some of the pressures that were on them. They quote from David Lange at the Lopdell Centre, talking about the fact that he saw ABLE as being set up to find ways of comparing the effectiveness of different schools. He was convinced that they were going to be able to find some "acute methodology", to enable these scientifically valid comparisons to be made. So far, we don't find ourselves overly convinced on matters value-added and I guess the PPTA is inclined to agree with ABLE that we don't see any evidence that this "acute methodology", has been found anywhere in the world yet.

There are a couple of strengths to ABLE which I'd like to comment on. The first is that it has recognised the problematic nature of assessment - that no form of assessment is value-free, you have to know why you're doing it, reorganise the potential for curriculum wash-back effects and that even if you have gained the information in an appropriate sort of way that it can be used inappropriately.

A second strength, especially considering what the system has been through in recent years, is the fact that they've tried to identify what's happening now and build on recent developments where they're seen to be able to serve the purpose. This is pretty important because you have to wonder how many more new ideas the system can handle at the moment. Certainly that is what teachers are saying about it.

There's one major proposal within ABLE that I want to say a few words about. It has something to do with the gap between the ideal and the real and it is, of course, to do with the issue of fourth form monitoring. I was relieved to hear Lockwood say that he was not particularly interested in it.

PPTA accepts that there could be merit in getting a picture of performance nationwide in key areas, on a regular basis. But we're not so sure about its being done for every fourth form student, in every school, every year, because there are some difficulties. Once it's every student, naturally they and their parents are going to want the results and that's totally understandable. (Although even that has its problems. I well remember a third form girl coming knocking frantically on my door and calling me out of the sixth form class and saying "Ms Hearn I'm really dumb." It was because her Maths teacher had told her an incredibly low raw score in a *PAT* Maths test. There was probably a purpose of the *PAT* test, but it wasn't served by that girl deciding that she was really dumb at Maths.) So, even if we accepted that parents and students get the information, how, under the Freedom of Information Act, do we stop it there? How can we stop the simplistic league table being published in the Auckland Star and then how can we stop people from drawing all kinds of inappropriate conclusions from them, about the nature of the different schools that are attended by the students who got particular test results?

Another worry is the danger of curriculum wash-back. System monitoring's a good idea if you can do it. If you can find ways of testing for functional literacy and numeracy skills without destroying the breadth of the curriculum - fine. But I can just see myself as a Head of English, coming under pressure to spend all our time teaching kids to read instruction manuals and fill in tax forms and write business letters. Of course, you don't learn to read instruction manuals by reading nothing but instruction manuals! It actually helps if you read some novels and poems and plays but not every principal, not every Board of Trustees and not every set of parents, is going to appreciate that.

The third worry is that if those two previous worries do eventuate, there's only going to be one year left in secondary school that's not dominated by some form of national assessment requirement.

The fourth form year has often been the time for some really creative teaching. You get the kids settled in in the third form and then in the fourth form you can do the extension work and the outdoor education and all kinds of integrated projects and activities that are about building skills and developing critical faculties and all those other things education is actually supposed to be all about.

The OECD examiners in 1983 commented that very few other systems had such a plethora of awards in the final three years of schooling. Now we've got rid of the University Entrance exam since then, but we're still left with one award in Form 5, one in Form 6 and two in Form 7 and if Form 4 testing did turn into just another exam, we'd be back to ages even darker than the one we had in 1983.

I accept that the ABLE Committee was aware of those dangers, but can't see the guarantees in their report that this scenario wouldn't come to pass.

Lockwood talked about his achievement initiative and in secondary schools there's no problem in saying that English, maths and science are going to be compulsory. They are already part of the core and providing there is still time for the other subjects, that's ok. But what makes me nervous is the immense confidence with which he and the National Party policy can talk about "clear learning objectives and attainment targets being established right through the core curriculum system from 5 to 15", and say that "people, progress and attainment will be monitored at key points against those attainment targets, using new assessment procedures". It seems to me that it sounds remarkably like David Lange's unfounded certainty that we could come up with an "acute methodology".

It's worth reminding Lockwood that so far we've had Merv Wellington's curriculum review and no conclusion, Russell Marshall's curriculum review - and that got turned into a research project on how to consult the community (which had some really interesting results but forgot about the original idea of what young people should learn and experience in school) and now apparently there is a group of Policy people in the Ministry beavering away, trying to come up with some national curriculum objectives which may be announced before the election. Who knows what's going to happen after that. The point I'm making is that it isn't as easy as it sounds, to set "clear learning objectives and attainment targets", much less assess them. I suppose it's just as well that teachers go on teaching something anyway, even though nobody's told them what the objectives are!

The comments that I've already made about the fourth form monitoring proposals, apply also to some of Lockwood's suggestions about testing and attainment targets. I acknowledge what he says about the positive and diagnostic purposes, in that it's designed to help, not hinder, but we still have a problem. If it's perceived as a national monitoring will it be for every student? Presumably so, if it's for diagnostic purposes. If so, who's going to get the information? Can it be used for inappropriate purposes? Are we going to have teachers spending all their time assessing and not enough time teaching?

Another question which applies both to Lockwood and to ABLE. Who would develop and administer the assessments? Inevitably there would be a need to create a new agency or to add to the staff of an existing agency.

I'm not one of those who goes in for recommending even further slashing of the bureaucracy - in fact I'd argue that one of the saddest aspects of the reforms has been the waste of valuable professional expertise from the system and the fact that there are now huge areas of institutional memory that have been completely lost (or that remain only in the bodies that are still more or less intact, like the PPTA!)

But I venture to suggest that there doesn't seem a lot of point in creating a bureaucracy to monitor national standards as a first priority, when there are huge black holes in several areas that are crucial to creating national standards in the first place. What I'm saying, in case it needs translating, is that before putting lots of money into national monitoring, there are at least three areas that need to be addressed.

Firstly, we need some national coherence in teacher training, instead of leaving it to chance with the individual teachers' colleges; secondly we need to restore a national system of curriculum development; and thirdly we need to look at the whole area of advisory services, teacher support and in-service training.

The reviews of advisory services and teacher training announced by the Minister are positive moves and I certainly hope that they're not going to be truncated suddenly, as a result of an election.

The other key area that I wanted to comment on is that of Lockwood's proposals for various sorts of awards in the senior school. Here I have to refer back to the OECD Examiners Report in 1983. Lockwood said that the purpose of assessment at secondary school was "to ascertain whether or not a student has met the requirements to receive various qualifications." The PPTA would argue that if we are now into the third phase of universal education, (in other words, if we now accept that the vast majority of students should be in some form of education through till age 18) then surely the goal in secondary school, at least till you're getting very near to the end, must be to keep them learning and keep them in the system as long as possible and keep the challenges going, so that they want to keep learning.

We can accept that when students are getting towards that final stage and particularly the final year, that it will be necessary to do what most systems do and have some national form of assessment with national moderation. But we don't see the point of doing the same thing at three different levels. If I understand Lockwood correctly, he's saying that in the fifth form he'll keep School Certificate, that he's going to keep Sixth Form Certificate, he's going to keep Bursary, he's going to return Scholarship and on top of that, he's going to add a new sixth and seventh form Vocational National Certificate. And the problem with that is that we'd just be back to the bad old days of U.E. We would have two awards at one level, three at the other, maybe four, if Higher School Certificate was still floating around. That presents us with several problems.

It has the obvious problem of the high status and the low status award. It doesn't matter what the intention is, it does not take kids long to spot which award is leading to university and so that's the one they and their parents want them to do, whether it's the most appropriate one for them to do or not. Yet a very significant part of the argument about U.E. was saying, let's have one award for sixth formers. This would take us right back to square one again.

The other question is how you staff it. It's always been an issue in the seventh form with Bursary and Scholarship. It's only been a handful of large schools, mainly boys', that have been able to staff Scholarships separately and that only in a handful of subjects.

If we were to return to having several awards at the one level, we would be back into the situation where only a small number of schools can staff them separately, where the majority are struggling and where smaller schools and rural schools are disadvantaged.

Interestingly though, I can comment on the sixth and seventh form notion that we could live with the idea of working towards some form of modular sixth and seventh award, if it were the only award. I have to say here that this is not developed to the point of being PPTA policy, but those of us who have been working in this area are very interested for instance, in developments like the Victorian Certificate of Education. We do see that the present situation is not well co-ordinated, with Sixth Form Certificate and Bursary as two separate awards and then the Link and Transition activities at times cutting across them. So we would be interested in exploring an approach which had a modular development through the sixth and seventh form, but only as a single co-ordinated development at both levels, not as a tack-on to existing awards.

On School Certificate we can understand the political reasons why ABLE chose to shelve it for a couple of years. The reality is that it can't disappear until it's been disentangled from Sixth Form Certificate moderation, but I cannot agree with Lockwood that it needs beefing up. The fact of the matter is that it was designed in the 1940's as a gate-keeping device, in the days when only a tiny elite went on to the final years of schooling. It was a ticket to the sixth form for them and a school leaving document for the rest. And today, the 69 percent who go on don't need it and it's not a helpful document for the ones who don't go on. And the more we talk about the need to increase participation rates, the more incongruous it is to have an award that acts as a disincentive right in the middle of a secondary schooling we actually want most students to complete.

On the issue of Scholarship I'll just say briefly that it's well documented that Scholarship has for many years served as an award for a very limited group of students and scholars, mainly boys, and a very narrow range of subjects, mainly maths and science and that the new single subject approach being taken in Bursary, looks a much better bet for recognising excellence subject by subject, rather than the dubious process of adding marks for different subjects.

And finally on internal assessment. Again it feels like going back - we've worked through all those arguments about school reputation affecting people's judgement about internal assessment. David Hood could do a better speech than me about the issue of educating employers and talking through those issues and building confidence - he and John Murdoch spent months doing exactly that after U.E. went. The fact is that internal assessment with decent national moderation is, if anything, more rigorous and can provide more useful information.

But I think we need to recognise that there has always been and continues to be, a tension between the desire on the one hand for national standards, the desire on the other hand for flexibility to meet the needs of individual students in the local schools. That's why it's particularly inappropriate to be talking national moderation, national standards, national demands, all the time when you're in the middle of secondary school. It's fair enough for a stronger emphasis on national moderation in the final year of schooling.

My last point is that if Lockwood was here, I'd be saying to him that we welcome his offer to work with people in education because we've had a fair while of being worked against. We've just got back into a more co-operative mode and we'd like it to continue. We think it needs to be recognised that teachers are at the heart of education and if teachers aren't involved in change, it won't work.

Ross St. George
Senior Lecturer in Psychology
Massey University

I want to begin by complimenting Shona Hearn (the previous speaker) on her very good presentation of many of the assessment issues we've been reviewing today. She's obviously far more familiar with some of the educational issues than I am now and she's properly and rightfully stolen some of my thunder on others.

A couple of quick points first in response to earlier discussions. I certainly had some disquiet about elements of the Project ABLE report when I first saw it, and I still do. Now, with the release of *Tomorrow's Schools* I certainly have some concern about the issues raised with respect to the retention of School Certificate. There is no future in transmogrification of School Certificate, if that is the right word, into some new award that is going to hang there still in the fifth form, like an albatross around the necks of many students.

Second, I'm not of the view that well-done, well-documented internal assessment cannot be pursued legitimately throughout this country, as it has been for many years. I think that the Opposition Spokesman on Education, should gather further information on these sorts of developments and the realities that people grapple with while implementing internal assessment.

Knowing that time's against us, I'm now going to turn to some more general comments on the ABLE Project and the ABLE documents. Some of these comments do touch on things that both the Minister and the Opposition Spokesperson have said about aspects of assessment.

When I received the first document I thought, "Well, those of us that are now starting to get enough grey hair will remember reading the Currie Commission Report, will remember reading Elley and Livingstone on internal assessment and external examinations and will remember also the Educational

Development Conference". Receipt of the ABLE report prompted me to scan through some of these documents. I could start pulling out themes that were the same and then I also remembered a little reddish book called *Educational Standards in the State Schools*. Bill Renwick and I locked horns over assessment matters at the time of this report. Our differences were over some of the ways in which data were collected, used and treated for particular purposes, namely political ends. Cynically, I would say these new green covered documents may suffer the same fate, that is, risk being ploughed under by the tide of political events.

That said, on studying *Tomorrow's Standards*, my views were, 'this reads better', 'it's more attuned to fostering student learning', 'it's good in parts.' Certainly it was better in many parts compared with the first ABLE document. But there are five themes that I do want to raise with you, to make comments on and to signal some cautions.

I think that many of the recommendations to do with the monitoring of schools, if done well, could be achieved. But I still question whether or not, as a country, we need to do this exercise. Do we not already have a good range of data sources about the nature of school outcomes and processes? I see a strong difference between the type of data proposed for collection and the type of data that stems directly from the assessment practices of teachers which can be used to good effect in the classroom. If the policy or set of recommendations of *Tomorrow's Standards* are pursued in full I do see, as others have already commented, that there is a danger of diverting a great deal of important teacher time away from the collection of assessment information that is vital to student learning and the teaching process in the classroom.

I am reminded of this from a period spent at Pittsburgh in 1986, where Cooley and Bickel were leading a project on 'School-based Monitoring of Achievement and Progress'. The schools received students from economically and socially depressed parts of Pittsburgh, with areas that could look like sets for 'Hill Street Blues.'

Their approach was to take good, well-rounded assessment methods and work closely with the teachers. The focus was on the daily monitoring of student learning - often it was do the basics, those important basics that are a foundation for the freedom to learn in other domains. The project used modern technologies (microcomputers) with teachers to implement important instruction flowing from the consequences of assessment within the hour, in the classroom. That's what teachers need to know. If kids are struggling with particular mathematics processes, particular constructions in language, particular concepts in introductory sciences and social sciences, geography, history, the time to know is now. If teachers and students know within the hour how they are doing, in fact because of the computer-based technology in far less than an hour using short multiple-choice type assessments, teachers could test, they could teach, they could re-test and the students sense their progress.

That's a far different cry from the appearance of outcomes that we might have to wait for three, six, nine months hence, when the kids have long moved through to the next topic or the next class or the next school and any learning difficulty has become some other teacher's problem.

The second theme that I would like to raise concerns the proposition that assessments that are unseen or private to teachers are in some sense superior. I think we're in danger of travelling down this track in New Zealand educational assessment and I've heard a couple of plea's along these lines today. It is implied that these private assessments would never be biased nor unfair and relate well to the things that are actually happening in the classroom. Conversely external assessment, often in the form of "standardized tests" are regarded as biased, unfair and unrelated to classroom learning.

I think that there is an important point which has to come through in any changes in the monitoring of student achievement and the methods of assessment, namely it has to be opened to public scrutiny. Currently a group of us are working on a project looking at real live teacher-generated assessment data, the sort of thing that gets put in students' record. It's disturbing the amount of information that is missing, often the information that would allow us to use that data for the constructive assessment of learning over time in the classrooms. That's the reality.

Why do teachers not record information critical to the meaningful use of their assessment data? Partly it's because of time, partly it's because of skill and it's also because of a lack of resources. They do not appear to document on-going current daily assessment tasks at all well and it would be frightening to leave them in that condition.

The third theme that I would want to comment on arising from *Tomorrow's Standards* is the emphasis on increasing the skills and understanding of educators in the measurement, assessment and monitoring fields which is to be applauded. Translating this into action or effect, however, to me is the critical issue. There's a difference between what I would call the investment in the tangible and the intangible, and it is one of the dilemmas we'll have to face. It's nice to have a scheme or ideas about this big business of national monitoring. Other countries are doing it, so we'll set up some kind of unit that has a location, name plate, people seen going in and out the door and it looks quite impressive. Occasionally people will visit schools. But the reality is, out there for most practising educators, they need help and assistance, guidance, professional upskilling if you like, in their schools. They're quite capable of doing the job, but I don't see that kind of assistance coming just from some Wellington-based dedicated monitoring unit.

The fourth theme. I don't think that we should be diverted, and I criticise the report at the moment on these grounds, from the issues that we face as a nation with respect to the educational achievement of identifiable sub-groups by labelling them as a result of assessment practices. I think we are in danger of going down that track, that is of arguing that somehow the outcomes, the differential achievement outcomes that we can monitor and we do see, are primarily consequences of assessment practices.

Quite clearly there is a diversity of educational goals that should be sought and different means by which they should be sought. This is not the issue. We have effects that are historical and there are contemporary effects that relate to differences in support systems that are external to the schools, that pupils bring and carry with them through their careers in education. Some of the consequences are manifest in educational achievement data outcomes.

I don't think we can label these outcomes, or carry all of these problems back to the door of assessment. Or, believe that in some form of modified assessment scheme, these would somehow be submerged, be hidden from view so that we can be publicly comfortable about the nature and the outcomes of education in New Zealand. That's not the kind of way to empower learners, to give them the skills and opportunities, to fully participate in this society or any other.

The fifth item that I want to raise from *Tomorrow's Standards* is one that others have commented on today. Cedric Hall in particular has noted the confusions in the use of terminology-like 'achievement-based assessment'. I've yet to see educational assessment that is not in some form, way or another, based on some kind of achievement.

On the criterion vs norm distinction, Gene Glass once said quite appropriately, of course, 'scratch a criterion and underneath you'll find the norm'. It was true when he said it, about ten years ago, it is true now. We've not taken on board the full meaning of his remark. Today I heard that we are also going to have AATs - Appropriate Assessment Tasks, whatever they might be, we're also going to have BAMs - Broad-based Assessment Measures, so I think we've got a lot of work to do with respect to clarifying our terminology. This is a necessary first step if we're going to move in any of the positive new directions signalled in *Tomorrow's Standards* for the assessment of pupils, schools and by implication, teachers.

In making some commentary on the original ABLE report I wondered whether the report might have had alternative title "DISABLE". I'm still worried that we might pursue assessment practices that result in Decisions Inhibiting Systematic Assessment for Better Learning. Hopefully implementation will prove me wrong.

Ken Wilson

Senior Advisory Officer, PPTA

Representing the Trade Union Education Authority

I had better explain the role I am assuming today. The Trade Union Education Authority have suggested to NZCER in the past that they would appreciate being invited to conferences of this nature and they were invited, but found themselves unable to be here. They acknowledged also that they had not had the opportunity to study Project ABLE in depth, so they approached PPTA and I undertook to talk with them and to bring their message. What I will say will be carrying a message from TUEA, who have a Trade Union perspective on ABLE. At the end I'll mesh some of the earlier comments with what I imagine they would have said, had they heard Lockwood speak to you today.

The first thing that they wanted me to say was to remind you that New Zealand is a classed society. There are numerous people out there who identify themselves as working class and they have a culture with goals and aspirations and that these are essentially democratic. Their expectations of the schooling system reflect the expectations they have for their capacity to live full lives in this country. They do not necessarily see schools as serving their interests and they are particularly cautious of any proposals on curriculum and assessment which do not adequately address the capacity of schools to differentiate children's achievement and life chances, on the basis of social class.

The other thing they wanted me to say was that the working class culture is not often recognised and taken into account and that they have equity concerns about that.

Similarly, they report great disappointment over nature and structure of the Boards of Trustees, where less than seven percent are people who may be readily identified as working class. The experience of working class people who had put themselves up for Boards of Trustees, was that their presentation of CVs and the nature and content of their CVs, did not compare with those of grey suits and grey suits got the jobs. So the nature of school Boards of Trustees, even in those areas where working-class children form the majority of the school population, are not representative of the community in the way that it was envisaged by Picot. That is serious and it is seen by TUEA as a concern.

Moving onto assessment, the view of the Trade Union Education Authority is that it recognises there are a variety of imperatives for this look at assessment and I think this has been confirmed this morning by Mr Goff and by Lockwood and other speakers, and that one of those imperatives was to address the crisis in education that had been constructed around 'falling standards', largely as a way of diverting attention from growing youth unemployment and addressing this issue through the performance of teachers and the failure of schooling. It's what others have called a 'structural policy response' - the government has moved to direct its resources towards managing attitudes, rather than addressing the actual problem. I think Lockwood's proposals highlight that.

The second point about assessment is that part of that policy has been to reconstruct the school curriculum and Marilyn Davies this morning gave a lucid outline of the room the Employers' Federation feels it has to alter and shape the school curriculum. There is considerable concern in TUEA at the nature of curriculum change and the narrowing of the curriculum to address economic goals. The virtues of a sustained and broad education are clear and valued by working class people and the possibility that we will have the Minolta Sixth Form Certificate or the Glaxo Seventh Form Science Project are regarded as unwelcome. What they're talking about really are the nature and functions of education, who runs schools, who is interested in School Certificate, where that debate is at and if the debate is to focus on assessment, they expect to have a voice.

A third point was that assessment functions in secondary schools differentially and has proved to be a barrier to working class children. School Certificate and University Entrance are good examples of that. I was staggered to hear Phil Goff suggest that New Zealand should select a comprehensive and rigorous assessment system. School Certificate has been extremely rigorous; it has kept half of the candidates out of continuing education for years.

Turning to some of Lockwood Smith's comments he noted that the removal of University Entrance has allowed more students to go through into the Sixth Form, where they have taken up a wider variety of courses, yet he seemed to be saying that in all of those, there will be an audit of those skills. That is not progress. The move is to open up the senior end of the secondary schools so that children will go through the fifth form, sixth form, and seventh form, but at Form 6 they may go to a polytech or some other provider. That picks up the virtue of a single certificate for the senior school. If Bursary or Scholarship type examinations remain at the levels Lockwood suggests, they can only have an elitist function in the nature of the skills they sample and in the link that those skills have with children's social and cultural origins. The children who come to school without the language or the skills are further penalised as these become a basis on which schools build.

The Trade Union Education Authority wishes to report on their experience in the Waikato, where working class children are now employed after school in adult jobs, because the breakdown of the National Award System and regional rates of pay make it possible to employ school children, where you might have had to employ a woman or a full-time worker. These children are working both to support their own families and to save for their own further education. Socially we're going back eighty or ninety years. The calls for further flexibility in the labour market, of course, will accentuate those trends.

The Trade Union Education Authority are mindful too of the ability of capital to reorganise schooling, so that the costs of education and skill-acquisition by workers are not borne by industry alone. I think that, in what Lockwood is saying, the needs of employers get a very strong emphasis, while the democratic needs and aspirations of working people don't get much at all.

Overall I think the Trade Union Authority would have said, having heard Lockwood, that Project ABLE as it lies is preferable, to his prescription for reform because it offers more opportunities and less disadvantages.

What Lockwood offers is really more whippings and a return to an attitude of meritocracy and performance, that I thought we had learnt our way through once and for all.

Finally Lockwood has twice now quoted Nash and Harker and Nash's review of the Smith and Tomlinson study. If he went to the next sentence, after the quotation he ascribes such significance to, he would find that it says, '... while we know now that schools can give rise to considerable differences in education progress, we have no real clue about the school practices which cause them!' The message from the Trade Union Education Authority is that in considering policy on assessment, social class must be accounted for.

Australian Developments in National Monitoring

Geoff Masters
Assistant Director (Measurement)
Australian Council for Educational Research

In Australia, as in many other parts of the world, there has been a great deal of activity recently to achieve national consensus on educational goals and to develop improved methods of monitoring student outcomes against these goals. Recent initiatives have been at two levels. At one level, there has been an attempt to achieve consensus among the Australian states and territories on goals for schools, to get systems working together to document ('map') current curriculum offerings in the various states and territories, and to agree on some general principles and frameworks within which all systems might work. At a second level, there has been a flurry of activity within state and territory education systems over the past eighteen months to introduce programmes for monitoring and reporting student achievements.

My plan today is to briefly describe some of these more recent developments on the other side of the Tasman. If there are lessons in any of this for what New Zealand might, or might not do in the area of national monitoring, then I will leave you to draw those lessons out of what I have to say.

The reasons for this recent activity in Australia are very similar to the motivations described by other presenters here today: demands for greater accountability of public institutions, concern over unnecessary duplication of effort, and claims that standards in schools have declined over time. In the Australian media there is a fairly constant barrage of criticism of standards of school achievement. This 1988 editorial from The Australian newspaper is an example:

> "Our education system is nothing short of disaster. There is literally nothing on the national agenda that cries out more urgently for attention than the need for major educational reform... The rot in Australian education starts with the schools. In education our aims have become confused and diluted. Our performance has slipped as a consequence and we do not even have the means to monitor our performance... Nothing short of a national crusade will be needed to transform the school system from the second-rate, slip-shod system of the present to something that answers the national and individual needs of Australians into the 21st Century."

That was followed by an article in the same newspaper entitled 'Why Our Education System is One of the Worst in the World' which claimed that 'we have no way of measuring the quality of education in Australia', we have 'one of the worst education systems in the developed world', and 'one with the least serious monitoring of its own standards'.

Of course, there is very little concrete evidence to support these claims. But there is almost no evidence to refute them either. Discussions of standards and international comparisons are carried out in Australia largely in the absence of evidence. As a result, discussions of standards are based largely on opinion and anecdote.

National Initiatives

We did conduct some national surveys of literacy and numeracy standards in 1975 and 1980. In those surveys, samples of students in all states and territories were tested in basic literacy and numeracy. The results of the 1975 testing generated controversy, mainly because of the way in which they were reported by the media. An article in the Bulletin magazine interpreted the results as evidence that standards had declined. The Director of ACER at the time, Dr Radford, wrote to the Bulletin pointing out that it is not possible to conclude on the basis of data collected at one point in time whether standards have declined or not. The Bulletin author's reply to Dr Radford was that we do not need evidence to know that standards have declined.

In part because of the way in which the results of the 1975 tests were interpreted by the media, there was concern among the Australian education community about the plan to repeat the exercise in 1980. In the state of Victoria, some schools boycotted the 1980 tests. The original plan to repeat this kind of national testing at regular intervals was shelved.

Despite boycotts in Victoria, ACER was able to analyse the 1980 data and to demonstrate that there had been no discernible decline in literacy and numeracy standards in Australian schools between 1975 and 1980. Ironically, this piece of evidence has been much used to refute claims of declining standards by those who were among the most vehement critics of the 1975 and 1980 testing programmes.

Within the last couple of years in Australia, as in most other western countries, there has been a new level of concern over our economic competitiveness. And, as elsewhere, the education system is being looked to as a way of providing the kind of workforce that we will need into the 21st century. The Australian Prime Minister has described the need for Australia to become a 'clever' country with greater numbers of people trained in mathematics, science and technology.

An influential document in recent discussions of educational reform in Australia was a 1988 paper put out by the Commonwealth Minister for Education entitled 'Strengthening Australia's Schools'. That paper called on the states to get together to develop a statement of national goals for schools and to try to reach agreement on benchmarks against which student outcomes might be monitored. The paper called for a national curriculum framework and a national approach to assessment.

Progress in achieving national consensus on matters of curriculum and assessment is slow in Australia because of our federal system and the fact that states have constitutional responsibility for education. Nevertheless, some progress has been made in the last year or two in achieving consensus in these areas.

In 1988 the Australian Ministers for Education, meeting as the Australian Education Council, reached agreement on a statement of broad goals for Australian schools. This is the first time we have had such a statement for all Australian schools. Following the release of the statement of national goals, state and territory education systems have cooperated in projects aimed at mapping specific curriculum areas. The purpose of this mapping exercise is to provide a picture of the curriculum as it exists in various Australian education systems. This process is revealing a high degree of commonality among our states and territories. Building on to these curriculum 'mapping' exercises, attempts are now being made to reach agreement on general curriculum principles: statements of the kinds of learning to which we believe all Australian students should have access. The extent to which these cooperative ventures move us in the direction of a common curriculum framework and common approaches to monitoring achievement in Australian schools remains to be seen.

State and Territory Initiatives

In parallel with these cooperative activities, most states and territories have moved over the last two years to put in place programmes for monitoring standards in schools. The first cab off the rank was the state of Victoria. The Victorian Labour Government established the Victorian Achievement Studies in 1988. The Victorian Achievement Studies are a series of studies conducted by ACER in government, catholic and independent schools. Tests are to be administered to samples of students in Literacy and Numeracy (1988), Science (1990), Social Science (1992), and Literacy and Numeracy (1993). Students are tested at 10 and 14 years of age. In that way, comparisons can be made back to the performances of 10- and 14-year-old students in 1975 and 1980. Some of the literacy and numeracy items we used in 1988 were also used in 1975 and 1980, and before that, in an IEA study. Some of the science items used this year were used earlier in the IEA second international science study.

The New South Wales Basic Skills Testing Programme was introduced by that state's Liberal Government in 1989. In the first year of that programme ACER developed and administered tests in aspects of literacy and aspects of numeracy to some 54,000 Year 6 students. A small sample of several thousand Year 3 students were tested as part of a pilot study to investigate the feasibility of extending the testing programme to all Year 3 students in 1990. This year we have tested all Year 3 students and all Year 6 students in NSW government schools. Some independent schools have also participated in the basic skills testing programme this year.

At Year 6, students are tested in two aspects of literacy (reading and knowledge of conventions of written English) and three aspects of numeracy (number, measurement and space). Parents of all participating students receive a report via the school. This report shows students' results in the various aspects of the tests in a graphical form and describes students' achievement levels in words. Teachers receive more detailed reports showing which questions each student answered correctly, which were answered incorrectly, and which incorrect answers each student chose. Schools receive school-level reports and ACER publishes a state report summarising students' test results.

In the state of Western Australia, samples of students will be tested for the first time this year in the third, seventh and tenth years of school as part of that state's Monitoring Standards in Education Programme. Those tests are concentrating in the first instance on writing and mathematics and will be extended to include science in 1991. The Education Ministry in Western Australia has assembled a relatively large test development team and is looking at a programme of standards monitoring that will continue over a number of years.

Queensland is testing samples of students in the fifth, seventh and ninth years of school this year. Tests will be in literacy and numeracy and will include test questions from the earlier 1975 and 1980 Australian Studies in School Performance to enable that state to compare literacy and numeracy standards with performances on those earlier occasions. As in Western Australia, a test development team has been established to develop tests locally. ACER is working in a consultative role with those teams, panelling test materials, assisting in trial testing of items interstate, providing test analysis software, and assisting in the writing of reports.

In South Australia, this is the first year of data collection in the Writing and Reading Assessment Project (WRAP). Unlike the other Australian systems, South Australia is not currently using externally developed tests. Instead the WRAP project collects samples of student work in the sixth and tenth years of school and provides a public commentary on these samples. An interim report of their first round of data collection has just been released.

As this brief description shows, there has been a great deal of recent activity in Australia to establish standards monitoring programmes in schools. This is not to say that there have not been some activities of this kind already underway in Australia. Queensland, for example, has conducted a series of 'periodic surveys' of student achievement over many years. Tasmania, too, has had its own monitoring programme in place. Following the 1975 ASSP tests in Australia, Tasmania decided to continue testing of this kind in mathematics and reading and to extend it to all students. They have tested students on a four-year cycle since 1975. Similarly, the Northern Territory and Australian Capital Territory have monitored standards in schools, often using published tests like the TORCH Tests of Reading Comprehension. The Australian Capital Territory currently has a 'Green Paper' canvassing options for future standards monitoring in that system.

Comparing Approaches

The approaches to standards monitoring being taken in Australia vary from state to state. In some states, such as Victoria, only samples of students are tested every four or five years. In other states, such as New South Wales, the intention is not merely to collect system level information but to also provide feedback to parents and teachers and so all students in a given year level are tested.

In Western Australia, the intention is to make assessment tasks available to teachers so that they can use them to monitor the performances of their own students in relation to the rest of the state. In this sense, the Western Australian project is similar in nature to a project to develop and norm test materials for teachers. The norming process provides the system with state-level data. The approach being taken in Western Australia is similar to the 'Benchmarks' programme of the Toronto Board of Education.

Despite these differences across Australian systems, there are also clear commonalities. There is not a great deal of interest in Australia in making comparisons for comparison's sake. There appears to be very little interest in knowing how students stand in relation to students of a similar age in other countries. We have been involved in IEA studies in the past, but there are few people in education systems at the present time who are interested in international comparisons of this kind. Similarly, there is almost no interest in making state-by-state comparisons of the kind now being made in the United States. Nor is there interest in comparing the performances of students in different types of school systems, government, independent and catholic, for example.

There is agreement among all states that it is undesirable to make the results for individual schools available to the public. This sometimes comes as a surprise to visitors from the UK and California where the public release of school-level results is seen as an important component of their testing programmes. It would be quite possible in states like New South Wales and Tasmania for school-level results to be made available, and at the time the NSW project was being set up, there was a fear that basic skills tests would be used to establish league tables of schools. In fact, this has not happened, and steps have been taken in Australia to ensure that schools do not publish their results on statewide monitoring tests.

Another common thread running through the recently-established monitoring programmes is a move away from minimum competency tests. In the 1975 and 1980 ASSP tests, as in many of the minimum competency testing programmes that were popular in the United States in the 1970s, test questions were limited to low-level survival skills such as reading a bus timetable, reading the instructions on a medicine bottle, filling in a form, and making a simple purchase. Tests now being developed are not limited to minimal skills of this kind but are being extended to assess higher-order thinking skills as well. 'Basic skills' in these programmes are not being interpreted as minimal survival skills but as fundamental areas of student learning, such as reading, in which higher-order skills are also important.

Most Australian systems are developing sequences of achievement levels in terms of which students' test performances are reported. These sequences often resemble the ordered levels being developed for profile components in the English national curriculum or the achievement levels used as reporting frameworks in the US National Assessment of Educational Progress. They go by various names including 'bands', 'skill bands', 'levels', and 'achievement levels'. There is now a move through the Australasian Cooperative Assessment Programme - of which New Zealand is a member - to develop shared subject profiles in mathematics and English literacy which will provide a set of profile components and achievement levels as shared assessment and reporting frameworks.

Other common features of monitoring programmes in Australia arise as a result of interaction among test development staff. Tests for the two largest Australian states are developed by staff of ACER's Measurement Division. Those staff are also in close contact with staff developing test materials in other parts of the country - Western Australia and Queensland, for example. It is probably not surprising that the tests being developed to assess Year 3 students' skills in measurement in New South Wales have much in common with the tests being developed to assess Year 3 students' skills in measurement in Western Australia.

Looking Ahead

An important issue in these Australian monitoring programmes is the question of the relationship between system-level tests and local curricula. Proposals to implement standards monitoring programmes are inevitably accompanied by high levels of teacher concern that the tests will constrain and distort teaching and learning. The response of some systems to this concern has been to limit testing to small samples of students, to not report test results for individuals or for schools and to report only aggregated, system-level data. In programmes of this kind there is little incentive for teachers to change their teaching practices. In general, this strategy appears to work. The downside may be that these programmes have no influence at all. Evaluations of the 1975 and 1980 Australian Studies of Student Performance, for example, failed to find evidence that those studies had any influence (either positive or negative) on what happens in Australian classrooms.

In New South Wales, the issue of potential curriculum distortion is addressed by linking the test materials tightly to the New South Wales syllabus statements in mathematics and language. In mathematics, for example, test items are focussed on the three strands of the mathematics syllabus: Number, Measurement and Space. In language, test questions are designed to mirror classroom activities in NSW schools, such as 'editing' other children's writing. In developing the tests for the New South Wales Basic Skills Testing Programme, ACER staff work closely with staff in that state involved in curriculum development.

Looking ahead, I expect that systems will make increasing use of a wider variety of assessment tasks in their monitoring programmes in an attempt to overcome the limitations of more traditional paper and pencil tests. In 1988 in Victoria, we tested written expression as part of that state's survey of literacy standards. Scripts were collected from students and graded at ACER. This year, our science survey in Victoria includes a number of practical tasks that require students to manipulate apparatus and to carry out some practical investigation. Another section of the test contains a number of open-ended questions designed to get a handle on students' conceptions of aspects of science. We refer to this part of the test as "Tapping Students' Science Beliefs (TSSB)".

Some of the materials we are developing will be made available as commercial test materials in the future. Teachers will be able to use these instruments with their own students and to make comparisons of their performances with the performances of a wider group of students.

Another issue that we will continue to address into the future is the question of the fairness of the instruments that we are developing to particular groups of students. We currently seek input from Aboriginal and Torres Strait Islander groups during test development and have carried out special analyses of the performances of these students at the request of the Ethnic Affairs Commission of New South Wales. Other groups, such as the Language Testing Unit at the University of Melbourne, have carried out independent evaluations of these tests for language and cultural bias.

A related question concerns the role of language in mathematics testing. Most states now emphasize the use of mathematics for problem solving and encourage teachers to teach mathematics in the context of meaningful problems. To be consistent with this curriculum emphasis, assessment procedures should test mathematics skills in realistic contexts. But the difficulty here is that to set mathematics in a context, language is required. Students from non-English-speaking backgrounds are likely to be at a disadvantage if they have difficulty understanding the problem. Our approach has been to try to set mathematics tasks in context but to minimize language demands. It has to be said that we have been only partially successful in our attempts at this, and this is a problem that we are continuing to address.

In developing test materials, we are making an attempt to set the questions in realistic contexts through the use of specially-prepared 'stimulus material'. In 1989 we printed a 'newspaper', a 'sale catalogue' and an 'Aussie magazine' as stimulus material for the tests. We have printed other magazines for the same purpose this year. In developing these materials we have gone to some trouble and expense to make the test materials colourful and interesting to students. The feedback we have had from schools indicates that students appreciate this effort and find the tests fun to take.

In the test booklets, students record their answers in a variety of ways, including underlining errors in text, colouring in notes and coins to indicate amounts of money, drawing paths on maps, and colouring in drawings of objects. These are then machine scored and analysed. We are continuing to explore new ways of using machine scoring and I expect that we will be developing even more innovative uses of this technology in the future.

The reporting of test results is also an area that we are continuing to explore. A frequent criticism of proposed testing programmes in Australia has been that tests will do nothing more than identify 50 percent of students as failing. To make it difficult for people to draw wrong conclusions of this kind, we have used item response theory to construct new reporting scales that do not invite inappropriate inferences about pass/fail levels. By not reporting number-right scores, we do not invite teachers and parents to interpret scores of less than 20 out of 40 as 'fails'.

For each of the 170,000 students tested as part of the New South Wales Basic Skills Programme in 1989-90, a parent report was produced and mailed home from the student's school. The parent report provides a graphical display of the student's results in the five aspects of the tests; Number, Measurement and Space, Reading, Language. At the bottom of the report there is a description of the skills typical of students with a given profile. On the back of the parent report is a description of the complete set of skill band descriptions for the five aspects of the tests. These have been translated into sixteen different languages for parents from non-English speaking backgrounds.

Teachers in NSW receive an even more detailed report showing which questions a student answered correctly and which were answered incorrectly. In 1990, this report also shows which incorrect answers students gave. The teacher report for Numeracy shows the test items in difficulty order with the easiest questions at the bottom of the page and the hardest at the top. The identification numbers of the questions answered correctly are printed on the left of the page, and the questions answered incorrectly are shown on the right. Teachers are able to use this report to identify relatively easy questions that a student answered incorrectly.

Schools in NSW receive other reports giving breakdowns of test results by age, gender, language background, and aboriginality for their school and for the state as a whole. In addition to these reports, a published report (Profiles of Learning: the Basic Skills Testing Programme in New South Wales, 1989) has been released describing Year 3 and Year 6 students' test results for 1989.

In much of the work currently underway in Australia a central question concerns the extent to which standards monitoring programmes can serve several purposes simultaneously. The variety of models being explored should provide us with a better basis for keeping parents, teachers, principals, school systems and the general public informed about the state of learning in Australian schools.

Concluding Remarks

Ian D. Livingstone
Director
NZCER

Many of you may be familiar with the book *Teaching as a Subversive Activity*. In winding up this final session, I'd like to advance the slightly tongue-in-cheek notion of assessment as a subversive activity.

Those of you with very long memories may recall the Currie Commission of 1961, and its recommendations for a series of 'checkpoint' tests, as benchmarks against which to measure the progress of pupils in core areas of the curriculum. In other words, a monitoring system, much like what is being proposed today. After some discussion between the Director-General of Education, the Director of NZCER, and teachers' organisations the final outcome was not a series of compulsory, summative tests, to be used at the end of the year, but rather a series of formative measures, the *Progressive Achievement Tests*, (PATs) designed to empower classroom teachers, and to be used by them at their own discretion, near the beginning of the year. Incidentally, the very first PATs, in Reading Comprehension and Vocabulary, with their level scores (linked to illustrative examples of reading material of suitable difficulty), were prime examples of what are now being referred to as 'criterion-referenced' tests. And that was over 20 years ago*!*

Those responsible for the terms of reference of Project ABLE were faced with a rather similar dilemma. One of the original reasons for setting up Project ABLE was undoubtedly related to the 'political' desire to monitor the country's educational system, to use assessment outcomes as a measure of its effectiveness. But along the way a little gentle subversion took place, first to name the project itself *Assessment for Better Learning*, and then to widen the terms of reference to place the emphasis squarely upon the learner, the individual student in the classroom. This is a form of subversion which I must say I heartily endorse.

As I look at the final report of the Project ABLE committee there is much that I can applaud, and I certainly approve of the general directions in which it is moving us, with its insistence that the interests of children are paramount, and its emphasis on identifying and reporting educational progress and growth.

But I do have some concerns. I would like to see an even stronger link established between curriculum and assessment. In the best of all possible worlds, curriculum should drive assessment, and not vice versa. Assessment procedures should be tailored to the curriculum being taught, and be an integral part of it. I was glad to see this emphasis coming through in *Tomorrow's Standards* where it was lacking in the first discussion document.

I would also like to make a plea for more precise definition of important concepts such as validity, reliability and usability of assessment measures. In a nutshell, I believe an understanding of validity is the key here, but an unreliable assessment procedure cannot be valid, and therefore should not be used, because it is likely to be unfair to students. I believe the report could have distinguished the different uses of assessment, formative and summative, and clarified such terms as criterion-referenced, standards-referenced and so on. Nor do I see a clear appreciation of the relationship which exists between norm-referenced and criterion-referenced assessment, or an acknowledgement of the sheer complexity of issues of assessment bias, its grounding in particular situations, and its relation to fairness for all students.

I was intrigued to see a principle stated in the beginning that self-assessment was the appropriate starting point for assessment, but I did not find this idea worked through in the report. I'd like to see it explored further, and to hear exactly what self-assessment consists of, how a student can undertake this in the classroom, and how it can become a reliable basis for other forms of assessment.

The idea of educational indicators, as measures of the performance of the system, is also prominent in the report. I would prefer to make a distinction between descriptors and indicators, and would be more stringent in validating an indicator than a descriptor, and require an indicator to have a clear link with expected outcomes before placing much weight on it. An indicator is not simply a piece of information about a school or system.

I'm very concerned about the need for training in assessment for all those involved with the process, and applaud heartily the comments made in *Tomorrow's Standards* about upgrading pre-service and in-service training in our institutions, so that assessment is done well, both accurately and fairly, and reported in ways which are meaningful to students, parents and the community.

In our discussions today we have been introduced to some helpful analyses of the Project ABLE report. I was particularly interested in the concept of quality assurance versus quality control. The very useful point was made that quality assurance is a reflection of the quality of the teaching and the other resources which go into our classrooms. If we get that right, we may not need to spend so much money on quality control of the outcomes. The Treasury comment on the opportunity cost of assessment is also worth taking account of. Assessment can be very expensive, both in terms of money, and in terms of the professional time taken to carry it out (which is another way of saying 'money'). We need to think very carefully about the cost of bringing about new forms of national assessment, and how cost-effective they will be in the long run in providing useful information for teachers and policymakers.

Finally, I would like to conclude with a question. Does national monitoring of standards (which I prefer to call levels of performance, reserving the term standards for an expectation of what should be, rather than what is) really bring about any changes in the system? The answer is, yes, it can. The United States results in the second mathematics study conducted by the International Association for the Evaluation of Educational Achievement (IEA) were sufficiently mediocre to bring about political action, a thorough investigation, and some educational change. In New Zealand, our third form results in the IEA mathematics tests also left something to be desired. This prompted the then Minister of Education, the Hon. M. Wellington, to request NZCER to take some initiatives in the preparation of materials for teachers of third form students. A series of self-diagnostic tests, with associated worksheets, was developed, known as the *Mathematics Topic Pre-tests*. Two tests, *Percentages* and *Integers*, are now available, and others on Measurement are in preparation.

So, yes, national assessment can bring about change in classrooms, but there does need to be a fast feedback loop between the monitoring itself, the public release of results, and the political processes which are able to do something about any deficiencies which are found. I have yet to be persuaded that we have the mechanisms in place to bring this about, and if the speed with which feedback of this sort can occur is too slow, it does not justify the very large expense involved. It may be that the better use of such amounts of money is to put more and better diagnostic measures into our classrooms, and empower teachers to use them more successfully. This could be a most cost-efficient use of scarce resources.

In conclusion, may I say that I am confident that our teachers will be able to subvert the bad and enhance the good features in the forthcoming reforms in assessment. I trust that this seminar will have helped to focus your thinking on just a few of the important issues involved. It is apparent from your reactions throughout proceedings that we have had a successful day. I wish to compliment the organizers and am pleased that NZCER has been able to make an early contribution to this important and continuing debate.

Contributors

Dr Terry Crooks, Director, Higher Education Development Centre, University of Otago.

The Hon. Phil Goff, M.P., Minister of Education.*

Carol Parker, President, New Zealand Educational Institute.

Dr Cedric Hall, Acting Director, University Teaching Development Centre, Victoria University of Wellington.

Marilyn Davies, Education Adviser, New Zealand Employers' Federation.

David Greig, Deputy Manager, Social Policy Branch, Treasury.

Katarina Mataira, Tutor, Waikato Polytechnic, Maori Language Commission.

Dr Lockwood Smith, M.P., Opposition Spokesperson on Education.**

Shona Hearn, President, Post Primary Teachers' Association.

Dr Ross St George, Senior Lecturer in Psychology, Massey University.

Ken Wilson, Senior Advisory Officer, Post Primary Teachers' Association.

Dr Geoff Masters, Assistant Director, Australian Council for Educational Research.

Ian Livingstone, Director, New Zealand Council for Educational Research.

---

* Subsequently, Mr Goff was not returned to Parliament.

** Subsequently, Dr Smith was appointed Minister of Education.

Conference Participants

Sandra Aitken
NZEI

Noeline Alcorn
Chairperson, NZCER Board

Ruth Anderson
Massey University

Teresa Ball
Palmerston North College of Education

David Burgon
Windley School

Jacky Burgon
NZCER

John Codd
Massey University

Con Coffey
New Zealand Principals' Federation

Cedric Croft
NZCER

Lesley Edgeley
N.Z. Qualifications Authority

Alison Gilmore
University of Canterbury

Bill Hamilton
NZEI

Clive Harper
Massey University

David Hood
N.Z. Qualifications Authority

James Irving
Ministry of Education

Angela Irwin
N.Z. Qualifications Authority

Colin Knight
Christchurch College of Education

Ian Livingstone
NZCER

John McLellan
Palmerston North College of Education

Tony McNaughton
NZCER Board Member

Ruth Mansell
Wellington College of Education

Ray Meldrum
N.Z. Qualifications Authority

Isobel Munro
Wellington Girls' College

Tony Naftel
Education Review Office

Roy Nash
Massey University

Jean Packman
New Zealand Principals' Federation

Margaret Palmer
Ministry of Education

Helen Patterson
N.Z. Qualifications Authority

Brian Pearl
Special Education Service

Val Podmore
NZCER

Jenny Poskitt
Massey University

David Philips
Education Review Office

Ted Preston
Education Review Office

Tom Reece
N.Z. Qualifications Authority

Amster Reedy
Education Review Office

Margaret Reeve
Christchurch College of Education

Neil Reid
NZCER

Les Robinson
Windley School

Margaret Rosemergy
Wellington College of Education

Dugald Scott
Wellington College of Education

Bob Stothart
Wellington College of Education

Margaret Stothart
Education Review Office

Ross Tasker
Wellington College of Education

June Vize
NZCER

Hans Wagemaker
Ministry of Education

Adrienne Waghorn
Wellington College of Education

Graham Wagner
NZCER

Janet Winter
Education Review Office

Conference on National Monitoring in Education

Programme:

9.00 a.m. Welcome

9.10 a.m. *Overview of National Monitoring in Education*
Terry Crooks
Chair - Valerie Podmore

9.50 a.m. *Government's Response to Tomorrow's Standards - Policies and Initiatives*
The Hon. Phil Goff, Minister of Education
Chair - Noeline Alcorn

10.30 a.m. Morning Tea

10.50 a.m. Symposium on Minister's Proposals

Participants	Nominated by
Carol Parker	NZEI
Cedric Hall	NZCER
Marilyn Davies	Employers' Federation

Chair - Cedric Croft

11.45 a.m. *Perspectives on National Monitoring*
David Greig (Treasury)

12.05 p.m. *National Monitoring in a Treaty of Waitangi Framework*
Katarina Mataira
Chair - Jacqueline Burgon

12.30 p.m. Lunch

1.30 p.m. *The Opposition's Response to Tomorrow's Standards - Policies and Initiatives*
Dr Lockwood Smith, Opposition Spokesperson on Education
Chair - Tony McNaughton

2.10 p.m. Symposium on Opposition Spokesperson's Presentation

Participants	Nominated by
Shona Hearn	PPTA
Ross St George	NZCER
Ken Wilson	Trade Union Education Authority

Chair - Neil Reid

3.05 p.m. Afternoon Tea

3.30 p.m. *Lessons From National Assessment in Australia*
Geoff Masters
Chair - Ian Livingstone

4.15 p.m. Close

Conference Organiser - Cedric Croft